Sussex Research Online

A parsimonious parametric model for generating margin requirements for futures

Article (Accepted Version)

Alexander, Carol, Kaeck, Andreas and Sumawong, Anannit (2019) A parsimonious parametric model for generating margin requirements for futures. European Journal of Operational Research, 273 (1). pp. 31-43. ISSN 0377-2217

This version is available from Sussex Research Online: http://sro.sussex.ac.uk/id/eprint/77556/

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:
Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

http://sro.sussex.ac.uk

A Parsimonious Parametric Model for Generating Margin Requirements for Futures*

Carol Alexander[a] **Andreas Kaeck**[b] **Anannit Sumawong**[c]

Abstract. Major exchanges employ the Standard Portfolio Analysis of Risk (SPAN) software to measure maintenance margins. However, its methodology has become cumbersome and opaque, having evolved over several decades and by now it requires that several hundred parameter values are re-set every day. We present a new, parsimonious parametric model for calculating margin requirements for futures which has a rigorous econometric foundation, being derived entirely from the median tail loss (MTL) of the returns distribution. This facilitates maximum likelihood volatility model calibration and state-of-the-art backtests. Then the parameters of the margin scheme which overlays the MTL may be calibrated using a variety of objectives. We examine three such objectives, including two which are designed to generate margins which mimic SPAN.

Key words: Finance, Backtesting, Margin Rules, Median Tail Loss, WTI Crude Oil

JEL: G12; G15; C53

---

*We would like to thank the Global Risk Institute (GRI) for funding this project and members of the GRI and the Bank of England for their insightful comments.

[a]**Corresponding author.** School of Business, Management and Economics, University of Sussex, Falmer, Brighton BN1 9RH, United Kingdom. E-mail: c.alexander@sussex.ac.uk

[b]School of Business, Management and Economics, University of Sussex, Falmer, Brighton BN1 9RH, United Kingdom. E-mail: a.kaeck@sussex.ac.uk

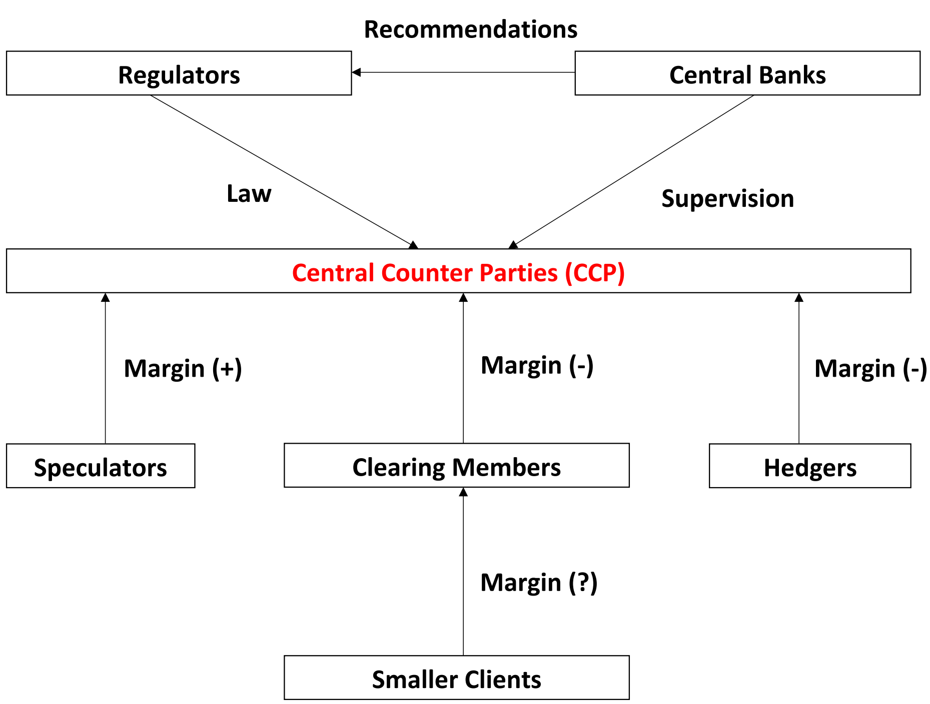
[c]School of Business, Management and Economics, University of Sussex, Falmer, Brighton BN1 9RH, United Kingdom. E-mail: a.sumawong@sussex.ac.uk

1 Introduction

Central counterparties (CCPs), also termed clearing houses, are regulated intermediaries which guarantee obligations and protect their members via effective default procedures. They aim to run a matched book but counterparty default can still expose them to considerable loss. They manage this counterparty credit risk in several ways, and in particular by taking collateral from participants in the form of margin payments. Speculators pay higher initial margins than clearing members and large hedgers, for whom the initial margin is usually set equal to the maintenance margin. Smaller clients of clearing members pay a margin which is determined by the member. When the margin account falls below the maintenance level a variation margin payment brings the margin back to the maintenance level. If a variation margin is not paid then a default identification process ensues. Figure 1 depicts the interaction between different parties involved with the CCP for setting futures margins.

Figure 1: **Margining in the Futures Market.**

Depiction of the margining setting process for the main players in the futures markets: (+/-) indicates that large speculators pay higher initial margins than clearing members and large hedgers, for whom the initial margin is set equal to the maintenance margin; and (?) indicates that smaller clients pay a margin which is determined by their clearing member. When the margin account falls below the maintenance level a variation margin payment brings the margin back to the maintenance level. If a variation margin is not paid then a default identification process ensues.



In September 2009 the G20 countries mandated a reform of transactions for over-the-counter (OTC) derivatives so that standardized contracts must be cleared by CCPs. In the USA, clearing houses must now adhere to the 2010 Dodd-Frank Act,[1] Title VIII of which reinforces the role of supervision over CCPs, in particular by the Securities Exchange Commission and the Commodity Futures Trading Commission. The Bank for International Settlements estimated that 62% of $544 trillion notional outstanding on OTC contracts was being cleared by CCPs by the end of 2016.[2] Thus, it has become very important that CCPs face the challenge of setting margin requirements which are risk sensitive enough to sustain minimal losses when liquidating defaulting positions, but also small and stable enough to remain competitive.[3] The tightening of regulations, combined with new margining for previously-OTC derivatives, presents a challenge for exchanges throughout the G20, not only in the US. For instance, the most recent European Market Infrastructure Regulation (EMIR) requires CCPs to maintain stable margins on all the products cleared, and the margin setting should be based on a risk model that is regularly back-tested to ensure that it provides a coverage level of at least 99% of the daily returns distribution.[4]

There are several reasons for these regulatory moves. More risk-sensitive margin requirements may be a means to reduce incentives for excessive speculation, driving the de-stabilizing speculators out of the market and thereby reducing volatility. Also, the recent wave of CCP mergers coupled with the continued expansion of derivatives markets raises the question whether CCPs may have positions so large that they have become 'too big to fail'. Counterparty credit risk can be highly contagious; indeed, some CCPs have now been designated as systemically important by the Financial Stability Oversight Council. Managing the risks of clearing houses in an event of a members' default is therefore of vital importance to reducing systemic risk in the world economy.

In 1988 the Chicago Mercantile Exchange (CME) group introduced the Standard Portfolio Analysis of Risk (SPAN) software, which generates price scenarios and takes the greatest loss as the margin. Margins aim to cover price movements with 99% coverage. The underlying risk metric is Value-at-Risk (VaR) or expected tail loss (ETL), also commonly called expected shortfall (ES). But no exact specifications are given about the risk model employed or the historical sample used for computation.[5] The methodology has evolved significantly over time, making it impossible to back-test properly. For example, old CME technical documents impose a margin coverage level of between 95-99%, whereas the current document states that all margins must cover at least the 99% Value-at-Risk. The time

[1] The Dodd-Frank Wall Street Reform and Consumer Protection Act Pub.L. 111203, H.R. 4173.

[2] See the latest report on www.bis.org/publ/otc_hy1611.htm.

[3] Here, and in the following, the term 'risk sensitive' means that the risk measure responds immediately to significant price changes. Note that volatility metrics derived from equally-weighted averages of squared returns are not risk sensitive because they suffer from 'ghost features' of an extreme event which may have happened far in the past but, due to the equal weighting, will appear just as important in the metric as if it happened yesterday.

[4] "A CCP shall assess its margin coverage by performing an ex-post comparison of observed outcomes with expected outcomes derived from the use of margin models. Such back-testing analysis shall be performed each day in order to evaluate whether there are any testing exceptions to margin coverage." Article 49 of EMIR (European Union, 2013). BIS (2015) establish principles for margin requirements for non-centrally cleared derivatives.

[5] However, the CME group did publish technical documentation in 2010, stating that SPAN employed several statistical techniques to capture distribution of price movements, depending on the product: e.g. fitting a normal-mixture distribution; fitting an extreme value distribution; and using exponentially weighted moving average (EWMA) volatility.

at which this change occurred is not mentioned in any current document. Given this lack of clarity, historical SPAN margins cannot be replicated accurately; one can only recreate the margins that would have resulted from applying the current methodology with the historical risk parameter files available.[6] Moreover, the evolution of the methodology has become very cumbersome and SPAN now requires re-setting hundreds of parameters every day.

This paper develops a parsimonious, intuitive, risk-sensitive model for setting the maintenance margin for futures contracts which could be used to replace SPAN. It's few parameters can be calibrated to replicate historical values of SPAN, but they need not be set so that margins mimic SPAN. In fact, the behaviour of our margin depends on the calibration objective and this is a matter of subjective choice. We propose three different calibration objectives, but others are possible within our framework. The model consists of two parts: a risk model and a margin model, each part having its own set of parameters, calibrated separately. Our empirical study implements the model for futures contracts on WTI crude oil, the S&P 500 index, the Euro-US Dollar exchange rate, and gold. We consider different asset classes to ensure that our margin model is not too specific to a particular market and may be adopted more broadly. It is well known that different asset classes have different features (e.g. exchange rates have more symmetric but leptokurtic distributions than equities, but equities are more skewed) and that different risk models may perform better for different assets, at different times.

We make the following original contributions to the literature on risk-based margin setting:
(1) We allow for different margins for long and short positions. Given the strong asymmetric tail behaviour that is commonly observed in futures, where upside risk is typically less than downside risk, exchanges will find it useful to apply our model for counterparties with un-matched long and short positions;
(2) Similarly, we are the first to develop a parsimonious margin model which captures the term-structure of futures contracts. In crude oil, and other contracts with strong term-structure features, margins should be greater for the more volatile short-term contracts. Again, exchanges should find our model useful for counterparties with calendar spread positions.
(3) The basic risk metric underpinning the margin model is median tail loss (MTL) rather than VaR or ETL (ES). The reasons for this choice are explained in Section 3. However, the margin scheme can take any quantile risk metric as input and so it is amenable to any type of instrument, provided only that one has a quantile risk metric available over a historical period that is sufficiently long to allow the margin scheme parameters to be calibrated.
(4) Margins for a set maturity, e.g. 1-month futures, cannot be calibrated to time-series created by rolling contracts. Data are corrupted by the artefactual jumps in prices which typically occur at the roll and the margins thereby created do not correspond to a fixed maturity. We avoid this pitfall by calibrating models to tradable returns on synthetic, constant-maturity futures. We also employ state-of-the-art econometrics for back-testing models for long and short positions simultaneously.

We implement and test our margin model using four different futures contracts but, for lack of

[6] E.g. see http://www.cmegroup.com/confluence/display/pubspan/Loading+SPAN+Risk+Parameter+Files for details on the CME risk parameter files for SPAN.

space, we only present our (extensive) back-test results for (part of) the term structure of WTI crude oil futures. In the following: Section 2 reviews the literature on margins for derivatives; Section 3 defines the risk models that we consider and describes our back-testing procedure; Risk-based margin schemes are discussed in Section 4 and used to motivate our model, as well as different calibration objectives; An empirical study is presented in Section 5 and Section 6 concludes.

2 Literature Review

The two main schools of thought on setting margin requirements are: the *prudential approach* of Figlewski (1984), Gay *et al.* (1986) and Booth *et al.* (1997), which argues that the main purpose of margins is to cover the clearing house's loss when participants default; and the *efficient contract design* introduced by Brennan (1986), which examines how margins and price limits can be set to make the contracts self-enforcing. Following these, the extant literature concludes that the optimal margin level should be high enough to cover the default risk faced by the clearing house when taking over positions, but low and stable enough to limit investors' opportunity costs and maintain liquidity in the market. Margin requirements should not be so high as to deter investors. Indeed, the highly-competitive clearing houses of today may drive margins down, in a "race-to-the-bottom" (Hardouvelis and Theodossiou, 2002). Interoperability may be desired by regulators, but it is competition and not cooperation that is currently on the CCPs agenda.[7]

Risk-sensitive margins should increase at exactly the time when systemically important members of the clearing house are required to hold more risk capital. Moreover, margin increases may have the effect of increasing volatility, thus increasing margins even further. Early academic research which confirms this procyclicality of margin changes includes Telser (1981), Hsieh and Miller (1990) and Kupiec (1993). Hardouvelis (1990) is the only work to find margins are countercyclical while others such as Kumar *et al.* (1991), Day and Lewis (1997) and Phylaktis and Aristidou (2013) claim that margin changes do not exacerbate market volatility. The procyclicality of margins is also a main focus for central bank's research: see Mägerle and Nellen (2011), Nahai-Williamson *et al.* (2013), Adrian and Shin (2014), Murphy *et al.* (2014) and Murphy *et al.* (2016). The need for margin-setting to limit procyclicality while also retaining sufficient risk-sensitivity is now acknowledged by CCPs, and the survey by Heckinger *et al.* (2016) has highlighted members' liquidity risk as a central issue for further research.[8]

Equilibrium asset-pricing models provide a theoretical foundation for procyclicality. Brunnermeier and Pedersen (2009) propose a model where margin increases can reduce liquidity and increase volatility, which in turn increases margins further, and the series of knock-on effects can cause liquidity to

[7]See derivsource.com/articles/margin-efficiency-ccps-portfolio-margining-eligibility-quest-interoperability.

[8]It is important to distinguish between margins for leveraging stocks and margins for futures because these may have different effects on procyclicality. For stocks, the margin account is composed purely of the investor's equity/debt levels, both the initial and maintenance margins are set as gearing ratios denominated in percentages. Historically, stock margin levels have fluctuated wildly on a year-to-year basis; changing to/from 50 - 100% at times, see Largay and West (1973) and Eckardt and Rogoff (1976) for example. Such extreme movements are not possible for futures margins, since these are obligatory.

completely dry out. Garleanu and Pedersen (2011) explain the mechanism for this, whereby large increase in margins drives de-stabilizing speculators out of the market at a time when hedgers need to exit positions, due to funding constraints caused by the margin increase. Speculators exit positions in response to margin changes more quickly than hedgers do, so hedgers will suffer from a lack of liquidity. Also, when hedgers are typically long, speculators can withdraw liquidity thus causing prices to fall as hedgers are forced to exit.[9]

Margins may affect different futures markets in different ways, because the balance between speculators and hedgers varies. The empirical study in Chou *et al.* (2015) finds that day traders on the Taiwan Futures Exchange are less sensitive to market traders than institutional investors. Another empirical study by Daskalaki and Skiadopoulos (2016) compares the effects of margin changes on commodity futures, currencies, equity and interest rates. They find that commodity markets are most sensitive to the effects of margin changes on speculation. In some markets (metals and grains) margin increases can impair the risk transfer from hedgers to speculators, because the speculators are more easily driven out of the market by onerous margins.

Now we discuss the specific strand of the margin literature that our paper contributes to, i.e. the empirical literature on SPAN margins and the papers which advocate alternative approaches for margin setting. Kupiec (1994) mimics daily values for SPAN, from December 1988 to December 1992, for calendar spreads on S&P 500 Futures, finding a coverage level of less than 95%.[10] More recently, Abruzzo and Park (2016) provide a detailed account of how SPAN margins have been changed by both the CME and ICE exchanges. Although a 'race-to-the-bottom' attitude, accompanied by chances to boost liquidity, should encourage CCPs to be prompt when decreasing margins, Abruzzo and Park (2016) show that the CME's decreases in margin requirements are more cautious than increases, and that the strain of competition forces the CME and ICE to alter margins in attempts to out-price each other. Daskalaki and Skiadopoulos (2016) examine the effect of large changes in SPAN margins, finding that large margin increases tend to decrease returns, and in certain markets (grains and metals) large margin increases impair the risk transfer mechanism because hedgers exit the market, being unable to find a speculative counterparty willing to share the risk. Large margin increases drive speculators who provide liquidity out of the market, causing an increase in the volatility of the futures contract.

Relatively few papers have attempted to develop a model for setting margin requirements which simultaneously satisfies the needs of regulators, CCPs and their clients. A simple rule is to set the margin exactly equal to the risk metric, as implemented in early works on margin requirements such as Booth *et al.* (1997), Broussard and Booth (1998), Longin (1999), Cotter (2001) and Lam *et al.* (2004). But this type of rule is too simple: it produces excessively volatile margins that exacerbate procyclicality because they change each time the risk is re-assessed. Such rules would be impossible

[9] By contrast, in the partial equilibrium model of Rytchkov (2016), which analyses the effect of time-varying margins on portfolio choice, the hedging demand produced by margin jumps should be comparatively small, and much less than that associated with price jumps. Other equilibrium asset-pricing models, such as Duffie and Zhu (2011) and Gibson and Murawski (2013), focus on the sub-optimality of margins and the potential benefits of interoperability within CCPs.

[10] By contrast, Kupiec and White (1996) tests SPAN against regulation T, a strategy-based system for margins on equities and equity options. For simple portfolios of S&P vanilla options, they find that SPAN yields lower margins than regulation T and provides 99–100% coverage of daily return movements.

to implement in practice without large costs to both the clients and the CCP.

Some papers have considered risk-based margin rules which have the type of buffer that is now recommended in EMIR, Article 28, as a means of reducing procyclicality.[11] The use of such a buffer produces much more stable margins and it is these papers that are very close to our research. Chiu *et al.* (2006) compares SPAN, as employed by the clearing house of the Taiwan Futures Exchange, with models based on a similar margin-change rule, but where VaR is the risk metric.[12] Extending this idea, Lam *et al.* (2010) also advocate a buffer, but instead of leaving the margin unchanged until the VaR increases or decreases by 15%, their margin scheme has parameters which can be calibrated to control the coverage probability and frequency of margin changes. Our empirical study will implement the margins proposed by both Chiu *et al.* (2006) and Lam *et al.* (2010), and compare the results with the CME SPAN margins as well the margins produced using our approach.

3 Risk Model Selection

3.1 Risk Metrics and Risk Models

For $0 < q < 1$ the $100q\%$ daily VaR is a loss which we anticipate exceeding $100q\%$ of the time, assuming the portfolio is not rebalanced over the next day. For instance, the 1% daily VaR is $-1\times$ the 1% quantile of a daily returns distribution, multiplied by the current price. Any model which can estimate VaR can also be used to estimate ETL, which is the average of the losses greater than the VaR. Both metrics are commonly used for regulatory capital calculations for market risk, internal economic capital allocation, and portfolio management. A comprehensive review of their estimation procedures can be found in Nadarajah *et al.* (2014) and Nieto and Ruiz (2016), respectively.

Because margin requirements should be set prudently we prefer to use either the mean or the median of the losses in the tail (i.e. ETL or MTL) rather than VaR as the risk metric for the model. Building on a recent and growing literature, we argue in favour of MTL as ETL is more affected by outliers and more difficult to estimate than MTL. Moreover, ETL is only jointly elicitable with VaR so it requires more complex back-tests than MTL, see Gneiting (2011) and Ziegel (2016).[13] Although MTL may fail on one of the properties of a 'coherent' risk metric, i.e. sub-additivity, Daníelsson *et al.* (2013) show that lack of sub-additivity is highly unlikely when parametric models are employed and Kou *et al.* (2013) show that MTL is less likely to be sub-additive than VaR. Moreover, Kou *et al.* (2013) criticize the sub-additivity axiom from the viewpoints of diversification and bankruptcy protection and argue that MTL is a superior risk metric because it falls into their class of *natural*

[11] "A CCP shall ensure that its policy for selecting and revising the confidence interval, the liquidation period and the lookback period deliver forward looking, stable and prudent margin requirements that limit procyclicality to the extent that the soundness and financial security of the CCP is not negatively affected. This shall include avoiding when possible disruptive or big step changes in margin requirements and establishing transparent and predictable procedures for adjusting margin requirements in response to changing market conditions. In doing so, the CCP shall employ at least one of the following options: (a) applying a margin buffer at least equal to 25% of the calculated margins which it allows to be temporarily exhausted in periods where calculated margin requirements are rising significantly;..."

[12] http://www.taifex.com.tw/eng/eng5/LSIndexFutComMargin.asp lists some details of the TAIFEX margin model.

[13] The MTL is easily elicitable, because it is a quantile.[14]

risk statistics.[15] Within this class they advocate MTL because it is the most *robust* to small changes in the data, especially in the tail.[16] Kou and Peng (2014a) claim that the tractability, elicitability, robustness and utility-foundation are three good reasons to prefer MTL to ETL.

Asymmetric tail behavior is typical of futures contracts. Upside risks are frequently lower than downside risks, in which case the tail above a $(1-q)$-quantile has lower probability mass than the tail below the q-quantile, for small q. Thus, our model should have the flexibility to specify different margins for un-matched long and short positions. Another reason why we choose the MTL metric is that it is suitable for capturing asymmetric proportional margin changes for up and down moves; this feature is inherent in MTL because volatility tends to increase sharply but decrease slowly.

To obtain parametric MTL estimates, we shall assume returns follow a dynamic process with either normal or Student t innovations, having cumulative density functions denoted $\Phi[\cdot]$ and $T_\nu[\cdot]$, respectively, where ν denotes the degree-of-freedom parameter. The corresponding $100q\%$ daily MTL, represented as a percentage of a futures price $F_{t,T}$ is, respectively:

$$\mathrm{MTL}^q_{t,T} = -\Phi^{-1}\left[\frac{\mathbb{1}_{q>0.5}+q}{2}\right]\hat{\sigma}_{t,T}\ , \tag{1}$$

$$\mathrm{MTL}^q_{t,T} = -T_\nu^{-1}\left[\frac{\mathbb{1}_{q>0.5}+q}{2}\right]\hat{\sigma}_{t,T}\sqrt{(\nu-2)/\nu}\ , \tag{2}$$

where $\hat{\sigma}_{t,T}$ is the time-t volatility estimate of the futures return $r_{t,T}=(F_{t,T}-F_{t-1,T})/F_{t-1,T}$.

We consider a number of univariate volatility estimation methods. The Exponentially Weighted Moving Average (EWMA), popularized by JP Morgan, is a particularly interesting model because it is used by the CME in SPAN. The variance estimate at time t is $\hat{\sigma}^2_{t,T}=\lambda\hat{\sigma}^2_{t-1,T}+(1-\lambda)r^2_{t-1,T}$. The smoothing constant λ cannot be calibrated but is rather set in an ad-hoc fashion. For instance, RiskMetrics^{TM} sets $\lambda=0.94$ for daily VaR estimates and $\lambda=0.97$ for monthly VaR estimates. For comparison, we also include a constant volatility based on a standard deviation estimated over a rolling sample of some fixed size N. To avoid the proliferation of qualitatively similar results in our empirical study, we set N to be either 30 or 90-days in the constant-volatility model and, in the EWMA models, we only present results for $\lambda=0.94, 0.96$ and 0.99.

The academic literature is dominated by models from the generalized autoregressive conditional heteroscedasticity (GARCH) family and we consider the standard symmetric GARCH, the GJR-GARCH model of Glosten *et al.* (1993) and Nelson (1990)'s exponential GARCH model.[17] The

[15] *Natural* risk statistics are characterised by different axioms to coherent metrics.

[16] In fact, Kou *et al.* (2013) investigate the robustness of MTL in considerable detail, according to four different statistics – see their Appendix F. Kou and Peng (2014b) prove that MTL is the only risk measure that (i) captures tail risk; (ii) is elicitable; and (iii) has a decision-theoretic foundation (i.e. it corresponds to the Choquet expected utility).

[17] GJR-GARCH models are given by $\hat{\sigma}^2_{t,T}=\hat{\beta}_0+\hat{\beta}_1 r^2_{t-1,T}+\hat{\beta}_2\hat{\sigma}^2_{t-1,T}+\hat{\beta}_3\mathbb{1}_{r_{t-1,T}<0}r^2_{t-1,T}$, and Nelson (1990)'s exponential GARCH model is $\ln\hat{\sigma}_{t,T}=\hat{\beta}_0+g(r_{t-1,T})+\hat{\beta}_3\ln\hat{\sigma}_{t-1,T}$, where $g(r_{t-1,T})=\hat{\beta}_1 r_{t-1,T}+\hat{\beta}_2(|r_{t-1,T}|-E[|r_{t,T}|])$. For instance Su *et al.* (2011) argue that GJR-GARCH is best for forecasting one-day-ahead downside risk and Chen *et al.* (2012), propose the use of Laplace innovations with GJR-GARCH. Non-parametric VaR methods interpolate the quantile from an empirical distribution of historical returns and these methods are reviewed and compared in Alexander (2008). The most popular method in this category is Barone-Adesi *et al.* (2002)'s Filtered Historical Simulation model, which

GARCH model parameters are calibrated using maximum likelihood on a rolling sample, with parameters being re-calibrated every day.

A novel feature of our margins is that they reflect the term-structure features of futures contracts. To implement this we capture correlations along the term structure using a multivariate GARCH model. Thus, in addition to the 16 univariate models considered above we also apply 12 different orthogonal GARCH (O-GARCH) and EWMA (O-EWMA) models (Alexander, 2001). That is, we first apply principal components analysis to the covariance matrix of returns and then estimate the univariate GARCH and EWMA models above on the components. We only consider models with one or two principal components, because otherwise the model can pick up idiosyncratic variation in just one part of the term structure, which is then reflected in the volatility of all maturities. We restrict the orthogonal models with two components to normal innovations because the Student-t distribution is not a stable distribution.

The literature on risk forecasting is vast and we discuss additional model specifications in the robustness section. In particular, following Lam *et al.* (2010), in Section 5.3 we use implied volatility data for forecasting the MTL, with either Gaussian or Student t innovations. Despite the breadth of models, we have excluded two well-known risk model classes, viz.: unconditional parametric models based on extreme value theory (EVT) – for which the classic reference is Embrechts *et al.* (1999); and models based on high-frequency, intra-day data – in particular, the heterogeneous autoregressive realised volatility (RV) models introduced by Corsi (2009). Such models may prove interesting avenues for further research on margining because they have found some success in predicting volatility, quantiles and quantile-related risk metrics. In particular, heterogeneous autoregressive models for RV have been successfully applied to quantile predictions of exchange rates (Clements *et al.*, 2008), as well as to volatility prediction of S&P 500 and US Treasury bond futures (Corsi, 2009) and crude oil (Sèvi, 2014). A useful empirical comparison of RV, EVT and GARCH approaches to quantile predictions for four asset classes may be found in Louzis *et al.* (2014).

3.2 Back-Testing Methodology

EMIR laws require exchanges to publish regular back-testing reports – and, for the reasons noted in the previous section, we seek an MTL model which passes back-tests on both upper and lower tails simultaneously. Below, we report empirical results for two back-testing methodologies. First, we employ the interval conditional coverage tests developed in Christoffersen (1998), for 1% and 99% MTL simultaneously. To this end, we assign a return to one of three intervals:

$$(1)\ r_{t,T} < -\mathrm{MTL}_{t,T}^{0.01};\ (2) -\mathrm{MTL}_{t,T}^{0.01} < r_{t,T} < -\mathrm{MTL}_{t,T}^{0.99};\ \text{ and }\ (3)\ r_{t,T} > -\mathrm{MTL}_{t,T}^{0.99}.$$

retains the shape of the empirical distribution via bootstrapping, while increasing risk sensitivity using a conditional volatility model. However, this model is computationally complex and as such it is not possible to employ it with the state-of-the-art back-testing techniques that we apply. We would have to reduce the number of simulations, and it is when insufficient simulations are used that MTL fails to be sub-additive – see Daníelsson *et al.* (2013).

Thus, interval (1) contains the negative returns which produce extreme losses for long positions, interval (3) contains extreme positive returns (losses for short positions) and otherwise the return lies in interval (2).

Denote the number of observed returns in each interval as n_1, n_2, n_3 where $n = n_1 + n_2 + n_3$ is the number of observations in the sample. The number of returns in interval i which are immediately followed by a return in interval j is denoted n_{ij} and the expected proportion of returns in interval i is denoted p_i. Then the relevant conditional coverage statistic is $LR^{cc} = -2(L(\Pi^0) - L(\hat{\Pi}^1))$, where $L(\Pi^0) = \sum_{i=1}^{3} \log(p_i)$ and $L(\hat{\Pi}^1) = \sum_{i=1}^{3} \sum_{j=1}^{3} n_{ij} \log\left(\frac{n_{ij}}{n_i}\right)$. Christoffersen (1998) shows that $LR^{cc} \sim \chi^2_6$. To reduce the volume of empirical results we do not report them separately for the unconditional coverage and independence tests, because they are combined into the conditional coverage test, although these results are available (on request).

Our second back-test builds on the continuous ranked probability score (CRPS) proposed in Gneiting and Ranjan (2011) which may be defined as:

$$C_w(f, z) = 2 \int_0^1 \left(\mathbb{1}_{z \leq F^{-1}(\alpha)} - \alpha\right) \left(F^{-1}(\alpha) - z\right) w(\alpha)\, d\alpha, \tag{3}$$

where f is the forecasting density of the risk model, F its corresponding cumulative distribution function and z the realized outcome. Intuitively, for the case $w(\alpha) = 1$ the score equates the integral over the squared difference between the forecasting distribution and the distribution resulting from perfect foresight, i.e. where all probability mass is on the realized outcome z. Thus the score measures the accuracy of the prediction with smaller CRPS statistics indicating superior forecasting performance. An advantage of this variant of the CRPS is that the weight function can place more emphasis on specific areas of the forecasting density. This is crucial for margin models, where accurate modeling of both tails is important. Gneiting and Ranjan (2011) propose to use $w(\alpha) = \alpha(1 - \alpha)$ for testing the center of the distribution, $w(\alpha) = (2\alpha - 1)^2$ for the tails, $w(\alpha) = \alpha^2$ for the right tail and $w(\alpha) = (1 - \alpha)^2$ for the left tail. We employ the same weighting functions in our empirical study.

To test the relative performance of two competing models we use the average CRPS, i.e. $\overline{C}_w^f = \frac{1}{n} \sum_{t=1}^{n} C_w(f_t, r_{t,T})$ where $\{r_{1,T}, \ldots, r_{n,T}\}$ are the T-maturity futures returns used in the out-of-sample test and f_t is the forecast density for $r_{t,T}$, based on information up to time $t-1$. The predictions of the MTL model with density f^i are preferred over another model with density f^j if $\overline{C}_w^{f^i} < \overline{C}_w^{f^j}$. Set $d_{ij,t} = C_w(f_t^i, r_{t,T}) - C_w(f_t^j, r_{t,T})$ and $\hat{\sigma}_{ij}^2 = \frac{1}{n} \sum_{t=1}^{n} d_{ij,t}^2$. Then the test statistic is $t_{ij} = \sqrt{n}\left(\overline{C}_w^{f^i} - \overline{C}_w^{f^j}\right) \hat{\sigma}_{ij}^{-1}$ and, under the null hypothesis of equal model performance, this is asymptotically standard normal. For technical details we refer to Gneiting and Ranjan (2011).

In order to allow multi-model comparisons, we compare the CRPS scores using the model confidence set (MCS) of Hansen *et al.* (2011) to identify models with superior performance at a given level of confidence $(1 - \alpha^*)$. Typically, α^* is set at 10% and/or 25%. A desirable property of the MCS is that it acknowledges the informativeness of the data, that is, the more informative the data, the fewer models in the MCS. To economise on space, in the main paper we only report MCS results for all our CRPS tests. Additional empirical results, including pairwise t_{ij} statistics, are provided in the online appendix. By considering so many GARCH models (and several underlying assets) we have ensured that our proposed risk model remains within the model confidence set, for all the assets we have considered, when tested against a broad spectrum of potential models. The plethora of models used at the backtesting stage therefore serves the important purpose of ensuring that our tests are thorough and comprehensive, and it allows us to draw conclusions for different underlying assets.

4 A Scheme for Setting Margins

Chiu *et al.* (2006) set the margin $m_{t,T}$ at time t for a futures of maturity T as:

$$m_{t,T} = \begin{cases} m_{s,T}, & 0.85\, m_{s,T} < \mathrm{VaR}_{t,T} < 1.15\, m_{s,T}, \\ \mathrm{VaR}_{t,T} & \text{otherwise,} \end{cases} \tag{4}$$

where $\mathrm{VaR}_{t,T}$ is the daily 99.7% VaR forecast at close of day t and s is the time of the last margin re-set.[18] In other words, the margin remains unchanged until the VaR increases or decreases by 15%, or more. The empirical study in Chiu *et al.* (2006) applies this margin rule to the exchange's SPAN risk model, as well as EWMA and Student-t symmetric GARCH models, finding that the best coverage is provided by the margins based on the t-GARCH VaR.

Lam *et al.* (2010) advocate a scheme based on $\hat{\sigma}_{t,T}$, an estimate of the contracts returns volatility at time t, and calibrate a parameter a such that $a\, F_{t,T}\hat{\sigma}_{t,T}$ represents a quantile of a normal or Student-t distribution for price variations. Rather than $+/-$ 15%, now the margin band has width determined by another calibrated parameter, b. This way, the margin is specified as follows:

$$m_{t,T} = \begin{cases} m_{s,T} & (1-b)\, m_{s,T} < a\, F_{t,T}\, \hat{\sigma}_{t,T} < (1+b)\, m_{s,T}, \\ a\, F_{t,T}\, \hat{\sigma}_{t,T} & \text{otherwise,} \end{cases} \tag{5}$$

where $F_{t,T}$ is the time t closing futures price and s is the time of the last margin re-set. The parameters a, which controls the coverage and b, which controls the frequency of margin changes, are calibrated by targeting a coverage of 99% and an expected number of 6 margin changes per year. A variety of volatility forecasts are compared as inputs to the margin scheme, and the authors conclude that implied volatility is preferable to statistical models of volatility.

Motivated by the ideas proposed by Chiu *et al.* (2006) and Lam *et al.* (2010), our margin setting scheme focuses on maintaining stability and reducing procyclicality, while also remaining competitive and risk-sensitive. By contrast with Chiu *et al.* (2006) and Lam *et al.* (2010), we advocate different margins for long and short positions, which is necessary to reflect the observed asymmetric tails in the futures' returns distributions. The margin is denominated in value terms, so we generate a series of dollar MTL daily forecasts, denoted $\mathrm{MTL}_{t,T}$ below, by multiplying the % MTL forecasts described in

[18] The 99.7% quantile was chosen to represent 3 times the standard deviation of a normal distribution, and which was the SPAN coverage level selected by the TAIFEX at the time of writing. Since then, the TAIFEX has targeted a slightly more modest 99% coverage level for SPAN. We use the 1% VaR in our empirical implementation of this rule.

the previous section by the current futures price $F_{t,T}$. Again we denote by s the time of the previous margin change and use the notation a and b for the parameters controlling the level of margin re-sets,[19] as well as parameters τ_{high} and τ_{low}, which determine the timing of this change.

Our margin will remain unchanged provided it remains within a corridor $\{1, 1+c\}\,\text{MTL}_t$, and will only change if it remains outside this corridor for a specified number of days. EMIR's Article 28 addresses the potential for procyclicality and instability of margins that are determined by the 1% and 99% quantiles of a daily returns distribution, by introducing a 25% capital buffer.[20] Therefore, we constrain the buffer parameter c to be at least 0.25. If the margin moves out of this corridor we do not change the margin immediately; instead the margin increases only when it has been below the lower boundary $\text{MTL}_{t,T}$ every day for the last τ_{low} days; a decrease occurs only when the margin has been consistently above the upper boundary $(1+c)\,\text{MTL}_{t,T}$ for τ_{high} days. Thus, the margin $m_{t,T}$ at time t for a contract with maturity T is given by:

$$m_{t,T} = \begin{cases} (1+c)\,\text{MTL}_{t,T}\,(1-b) & \text{when } (1+c)\,\text{MTL}_{t,T} \leq m_{t-j,T}\ \forall\, j = 1, ..., \tau_{\text{high}}, \\ \text{MTL}_{t,T}\,(1+a) & \text{when } \text{MTL}_{t,T} \geq m_{t-j,T}\ \forall\, j = 1, ..., \tau_{\text{low}}, \\ m_{s,T}, & \text{otherwise}, \end{cases} \tag{6}$$

where s is the time of the last margin change. In the above, the MTL is a one-day-ahead forecast of 1% MTL for a long position and of 99% MTL for a short position.

Our empirical observation of MTL forecasts will demonstrate that they tend to jump upwards rapidly but decrease more slowly. This is a consequence of volatility clustering, which is a well-known feature of commodity, equity and currency markets. As a result, our margin increases tend to be proportionally larger than margin decreases. This type of asymmetry is also observed in CME margins for these products. To support this, Daskalaki and Skiadopoulos (2016) find that commodity futures tend to exhibit larger proportional margin increases than decreases. However, CME margins for 5-year Treasury note futures (which exhibit much less volatility clustering in their returns) have greater decreases than increases.

In summary, there is no increase (decrease) in the margin unless risk, as measured by MTL, has remained too high (too low) for a sufficiently long time. Larger tolerances imply more stable margins and different values for the high and low tolerance parameters may induce asymmetry in the frequency of margin increases relative to decreases. The CCP may prefer $\tau_{\text{high}} < \tau_{\text{low}}$, so that margin decreases are more frequent than increases. This type of asymmetry may or may not be observed in SPAN. For instance, in their historical examination of CME margins, Abruzzo and Park (2016) find that some margins (e.g. currency futures) move downwards more frequently than upwards; but the opposite is the case in other markets (e.g. agricultural futures) and Daskalaki and Skiadopoulos (2016) also find that precious metals have more frequent margin increases than decreases. This evidence that CCPs

[19]This is to facilitate presentation of our results in the same table as the results of applying the scheme introduced by Lam *et al.* (2010). But note that the parameters a and b have different meanings in each scheme.

[20]The Article (see footnote 7) explains how the buffer can be used, but no recommendations are given regarding how one should replenish it. Our parameters a and b specify how the buffer is replenished.

set margins in different ways, depending on the type of futures contract, motivates these asymmetric tolerance parameters in Equation (6).

The parameters $\{a, b, c, \tau_{\text{high}}, \tau_{\text{low}}\}$ may be calibrated using a variety of objectives. Different parameters may be used for long and short positions, based on 1% and 99% daily MTL forecasts respectively. For instance, the CCP may wish to calibrate parameters so that the scheme produces margins similar to SPAN. Or the CCP may simply wish to make margins as stable as possible, because margin changes can deter investors and may have pro-cyclical effects on liquidity and volatility (Brunnermeier and Pedersen, 2009). Alternatively, they may set target statistics such as the mean and variance of the size of margin changes. Other targets could limit the range of possible changes, and other alternatives are to choose parameters to minimize the number, or the average size, of margin changes.

The features of our margin depend on the parameter values and these depend on the calibration objective employed. To illustrate this in Section 5.2 we provide an empirical comparison of margins based on (4), (5) and (6) when the parameters are calibrated to different markets using the following calibration objectives:

- E *RMSE*: Minimize the root mean square error (RMSE) between our margin and a target margin. This rule focuses on the CCPs competitiveness because when they have previously used SPAN it would be uncompetitive to set margins far away from SPAN levels. This objective may also be employed when one CCP seeks to track another's margin changes;

- T *Target Stats*: Specify targets such as the average number of margin changes, or their average size, and minimize the sum of squared standardized deviations. This more flexible rule allows the CCP to control the statistical features of margins and the targets may also be set relative to the statistical characteristics of a target margin;

- S *Stability*: Set parameters so that the resulting margin has as few changes as possible.

5 Empirical Study

The aim of this section is to select the best MTL model(s) and use them to construct daily time series for margins on futures contracts for West Texas Intermediate (WTI) crude oil, the S&P500 index, gold, and the EUR/USD exchange rate. The features of these margins are then examined and compared with the CME margins and those obtained using the margin schemes advocated by Chiu *et al.* (2006) and Lam *et al.* (2010).

For the sample from 31 December 2008 to 7 August 2016 daily closing prices for the futures contracts of all available maturities were downloaded from Bloomberg. From these we follow the methodology introduced by Galai (1979) to construct daily, investable, constant-maturity returns on the available term structure of futures prices. That is, we linearly interpolate between daily returns (not the prices) on the two contracts straddling a fixed maturity, taking the bid and offer prices into account when rebalancing the weights on each futures contract. This methodology is more complex than usual, but it ensures we measure margins on the appropriate maturity of risk and it avoids the artefactual jumps that can occur when historical time series are created by rolling futures contracts. Additionally, our time series really do represent realizable returns.[21]

The CME's SPAN risk parameter files are used to re-construct daily CME margins for all four underlyings from 2 January 2009 to 7 August 2016. This is based on the current methodology for constructing margins.[22] Being risk-based, our margins vary with term of the futures. But for gold, S&P 500 and Euro/Dollar futures we can only assume the CME term structure is flat and equal to the first-to-mature series, there being no term-structure in the CME margins for these products. Also, unlike ours, the CME margins are identical for long and short positions.

5.1 Back-Testing Risk Models

Our back-testing procedure is exhaustive, requiring the examination of numerous outputs on 28 different MTL models, each applied to several different types and maturities of futures. There is no space to report all the results here and we choose to focus on WTI crude oil futures because they have the most complete term structure data. The other contracts that we study are only traded on the quarterly cycle, but WTI contracts expire monthly and start active trading more than a year before expiry. Also, the qualitative conclusions that we draw for WTI are similar to those for the other three term structures. Detailed results for other futures contracts are available in the on-line appendix.

We further downloaded daily futures prices from 31 December 1989 to 31 December 2008 specifically for the testing of MTL models, prior to calibrating and testing of the margin scheme. Then we estimated all 28 MTL models for the investable, daily returns on 1-month, 2-month, ... 12-month synthetic constant-maturity contracts. It is sufficient to present results for the returns on 1-month, 3-month, 6-month and 12-month contracts, these being representative of the sample; and it is necessary to present only a sub-set of our results to keep this paper to an appropriate length.

Sample statistics are reported in Table 1. The largest negative return, which occurs during the Gulf War in January 1991, induces a large excess kurtosis, especially in the 1-month futures where a return of -31.75% was recorded on 17 January 1991. As is typical with energy futures, the returns standard deviation decreases with contract maturity and there is a negative skewness and high kurtosis, indicating a pronounced asymmetry in the tail weights of their returns distributions.

The 28 competing MTL models are first estimated on an in-sample period spanning the 1500 days ending on the last trading day in 1995. Then the sample is rolled forward day-by-day, keeping the sample size constant at 1500, and the GARCH model parameters are re-calibrated daily, the sample standard deviation is re-calculated for the unconditional volatility models, and the EWMA volatilities are updated. This way, for each model we produce time series of out-of-sample daily 1% and 99% MTL forecasts in the lower and upper tail of the futures return distributions, from 1 January 1996 to 31 December 2008. The back-tests described in Section 3 are performed on this sample. The later

[21] We require results for a constant maturity, not one with systematically varying term. If the term structure is flat, as it is for gold futures, there will be little difference between the two possibilities. But for crude oil, the roll cost can be considerable, as the term structure fluctuates between steep backwardation and strong contango. Then it is important to take the roll cost into account so that we analyse investable returns.

[22] See http://www.cmegroup.com/clearing/risk-management/historical-margins.html.

Table 1: **Summary Statistics for WTI Crude Oil Futures**

Summary statistics for daily returns on synthetic, constant-maturity WTI futures of 1-month, 3-months, 6-months and 12-months maturity from 2 January 1990 to 31 December 2008. In percentage points.

	WTI 1M	WTI 3M	WTI 6M	WTI 12M
Average daily return	0.05	0.06	0.05	0.05
Standard deviation	2.18	1.89	1.69	1.50
Skewness	-0.68	-0.57	-0.44	-0.21
Kurtosis	15.51	12.24	10.38	7.93
Min return	-31.75	-24.74	-19.27	-12.23
Max return	13.96	12.52	10.94	10.44

sample, starting in January 2009, is reserved for out-of-sample forecast evaluation.

Table 2: **Conditional Coverage Tests**

This table reports conditional coverage statistics for 1% and 99% daily MTL simultaneously, based on daily returns on 1-month, 3-month, 6-months and 12-months WTI crude oil futures. The back-testing sample is from 1 January 1996 to 31 December 2008. Rejection of the null hypothesis that the model provides adequate coverage at 5% or 1% is denoted * or ** respectively.

Model	1 months	3 months	6 months	12 months
Normal GARCH	10.68**	18.79**	19.39**	25.80**
t-GARCH	1.41	0.71	0.13	0.63
Normal EGARCH	6.84**	21.41**	24.28**	27.08**
t-EGARCH	0.33	0.01	0.84	1.46
Normal GJR	7.33**	20.20**	26.90**	23.45**
t-GJR	0.71	0.51	0.20	0.84
Normal EWMA 0.94	25.03**	23.45**	21.93**	16.65**
t-EWMA 0.94	25.03**	23.45**	21.93**	16.65**
Normal EWMA 0.96	14.17**	13.47**	17.72**	16.65**
t-EWMA 0.96	14.17**	13.47**	16.65**	16.44**
Normal EWMA 0.99	23.96**	25.43**	36.23**	15.38**
t-EWMA 0.99	19.86**	15.24**	16.44**	39.29**
Normal Constant 30	41.63**	44.75**	38.75**	26.84**
t-Constant 30	34.76**	36.89**	31.25**	42.49**
Normal Constant 90	23.45**	22.69**	13.91**	31.30**
t-Constant 90	20.46**	18.86**	14.80**	31.30**
Normal GARCH 1	29.83**	22.63**	34.20**	54.40**
t-GARCH 1	4.22	3.84	3.94	21.87**
Normal EGARCH 1	29.83**	32.62**	31.93**	54.40**
t-EGARCH 1	4.22	3.65	6.65	28.81**
Normal GJR 1	32.21**	27.60**	34.57**	55.02**
t-GJR 1	4.16	3.84	6.70	26.56**
Normal EWMA 0.94 1	41.71**	36.65**	38.70**	61.84**
t-EWMA 0.94 1	41.71**	36.65**	38.70**	61.84**
Normal GARCH 2	22.44**	21.64**	34.20**	42.31**
Normal EGARCH 2	21.14**	30.69**	30.57**	48.86**
Normal GJR 2	20.80**	27.60**	34.57**	45.34**
Normal EWMA 0.94 2	26.52**	32.41**	37.40**	48.27**

Table 2 reports the results of the Christoffersen (1998) conditional coverage tests on 1% and 99% MTL simultaneously. Again, to save space, we only report results for 1-month, 3-month, 6-month and 12-month futures (but make available on request the results for other maturities, and the corresponding unconditional coverage tests). The table shows that the three univariate GARCH models with Student-t innovations provide adequate coverage but all the other univariate models can be rejected at the 1% significance level. The three orthogonal GARCH models with Student-t innovations based on only one component also provide good coverage, except for the 12-month futures. The longer term WTI contracts exhibit idiosyncratic behaviour, possibly associated with less frequent trading, which is out of line with the common trend in shorter term contracts. None of the other multivariate models provide adequate coverage.

As our second back-testing methodology, we use the CRPS loss function and report the MCS p-values of Hansen *et al.* (2011). Results for WTI 1-month futures are in Table 3. The right-tail results in column 4 (Right Tail) show that all models are included in the 10% model confidence set. In fact, even the constant-volatility models (which perform so poorly in simultaneous coverage tests) remain in the 25% MCS and there is little to distinguish between the models for upper-tail prediction. For other choices of weights the constant-volatility models, the EWMA models with slow decay (i.e. $\lambda = 0.99$) and the normal symmetric GARCH do not perform well. Confirming the other back-tests, the best models are: the univariate t-GJR, (without tail weighting or when both tails are weighted); the t-EGARCH (in the right tail); and symmetric t-GARCH (in the left tail). Again, the multivariate GARCH models with one principal component lie in the 10% MCS, but the multivariate EWMA models are excluded. However, the EWMA models with $\lambda = 0.96$ lie in the 25% MCS with any tail weight. This is notable, given the simplicity of using EWMA models, relative to GARCH.

Results of other maturities are qualitatively similar to those in Table 3, except for 12-month futures for which the MCS results are presented Table 4. Here, the single-component O-GARCH models are now equally good at left tail predictions – they are even in the 25% MCS – and, remarkably, the univariate t-EWMA with $\lambda = 0.96$ is the best model for both tails, right tail and centre predictions. Since it is also in the 25% MCS for other maturities, the t-EWMA with $\lambda = 0.96$ is clearly an interesting model to investigate, because it is much easier to implement than a GARCH model.

To summarise the back-test results, the class of univariate t-GARCH models appears particularly well suited for simultaneous tail prediction, especially the t-GJR model. But GARCH models require regular parameter re-estimation, and so – for its simplicity – the univariate t-EWMA with $\lambda = 0.96$ may be preferred. This is not the best model but, using the Gneiting and Ranjan (2011) test provided in the internet appendix and the Hansen *et al.* (2011) MCS, we have found no strong statistical evidence of inferiority to the other models. Our back-testing results for S&P500 index, gold, and the EUR/USD exchange rate are comparable (untabulated). Therefore, the next section will compare the margins resulting from the t-GJR with those based on the t-EWMA with $\lambda = 0.96$.

Table 3: **Model Confidence Sets for MTL models: WTI 1-month futures.**
This table reports the p-values for Hansen *et al.* (2011) model confidence sets of MTL models estimated on daily returns on 1-month WTI crude oil futures. The loss function is the continuous ranked probability score (3) with different weight functions: $w(\alpha) = 1$ for no weight; $w(\alpha) = (1-\alpha)^2$ for the left tail; $w(\alpha) = \alpha^2$ for the right tail; and $w(\alpha) = (2\alpha - 1)^2$ for both tails. Test data sample 1 Jan 1996 to 31 Dec 2008. Models with $*$ and $**$ belong to the 10% and 25% MCS, respectively. A p-score of 1.0000 indicates the best model.

Model Name	No Weight	Left Tail	Right Tail	Tails
Normal GARCH	0.0251	0.2947**	0.1247*	0.1754*
t-GARCH	0.7930**	1.0000**	0.1649*	0.8608**
Normal EGARCH	0.0761	0.0164	0.7024**	0.4114**
t-EGARCH	0.7930**	0.3235**	1.0000**	0.8608**
Normal GJR	0.1232*	0.2529**	0.2130*	0.2077*
t-GJR	1.0000**	0.7951**	0.3467**	1.0000**
Normal EWMA 0.94	0.0761	0.3235**	0.1280*	0.2077*
t-EWMA 0.94	0.0761	0.3235**	0.1280*	0.2869**
Normal EWMA 0.96	0.4011**	0.3235**	0.6362**	0.8521**
t-EWMA 0.96	0.4146**	0.6702**	0.6981**	0.8608**
Normal EWMA 0.99	0.0116	0.0507	0.1280*	0.0343
t-EWMA 0.99	0.0251	0.1045*	0.1316*	0.0645
Normal Constant 30	0.0029	0.0047	0.1264*	0.0072
t-Constant 30	0.0039	0.0063	0.1280*	0.0107
Normal Constant 90	0.0055	0.0086	0.2286*	0.0186
t-Constant 90	0.0077	0.0240	0.2130*	0.0249
Normal GARCH 1	0.0075	0.2529**	0.1130*	0.0186
t-GARCH 1	0.4146**	0.7610**	0.1316*	0.7287**
Normal EGARCH 1	0.0251	0.0086	0.6702**	0.0343
t-EGARCH 1	0.2995**	0.2529**	0.4667**	0.3363**
Normal GJR 1	0.0251	0.2529**	0.1280*	0.0645
t-GJR 1	0.5590**	0.7951**	0.2286*	0.7990**
Normal EWMA 0.94 1	0.0052	0.0507	0.1247*	0.0158
t-EWMA 0.94 1	0.0052	0.1045*	0.1247*	0.0168
Normal GARCH 2	0.0052	0.1045*	0.1191*	0.0186
Normal EGARCH 2	0.0251	0.0086	0.7024**	0.1719*
Normal GJR 2	0.0251	0.2529**	0.1280*	0.1719*
Normal EWMA 0.94 2	0.0075	0.0507	0.1280*	0.0214

5.2 Calibrating Margin Parameters

Here we calibrate the margin setting scheme (6) to the MTL series generated by two volatility models selected in the previous section, i.e. (i) t-GJR and (ii) t-EWMA with $\lambda = 0.96$. Having used data from 31 December 1989 to 31 December 2008 to back-test the MTL model, our sample now covers the period 1 January 2009 to 7 August 2016 so that the first daily return coincides with the period for which CME margins are available. Our analysis is based on maintenance margins which, following Abruzzo and Park (2016) we standardize by the contract size. For example, the maintenance margin on S&P 500 futures is (at the time of writing) $25,000 and the point value of the futures is $250, so

Table 4: **Model Confidence Sets for MTL models: WTI 12-month futures.**
This table reports the p-values for Hansen *et al.* (2011) model confidence sets of MTL models estimated on daily returns on 12-month WTI crude oil futures. The loss function is the continuous ranked probability score (3) with different weight functions: $w(\alpha) = 1$ for no weight; $w(\alpha) = (1-\alpha)^2$ for the left tail; $w(\alpha) = \alpha^2$ for the right tail; and $w(\alpha) = (2\alpha - 1)^2$ for both tails. Test data sample 1 Jan 1996 to 31 Dec 2008. Models with $*$ and $**$ belong to the 10% and 25% MCS, respectively.

Model Name	No Weight	Left Tail	Right Tail	Tails
Normal GARCH	0.1706*	0.3430**	0.0002	0.0393
t-GARCH	0.8356**	1.0000**	0.0005	0.9864**
Normal EGARCH	0.5803**	0.0303	0.3586**	0.5981**
t-EGARCH	0.6593**	0.2519**	0.1401*	0.9557**
Normal GJR	0.2124*	0.1322*	0.0003	0.0714
t-GJR	0.7484**	0.4445**	0.0005	0.9557**
Normal EWMA 0.94	0.6593**	0.1322*	0.2252*	0.7360**
t-EWMA 0.94	0.6655**	0.1472*	0.1401*	0.7447**
Normal EWMA 0.96	0.8356**	0.0790	0.4629**	0.9557**
t-EWMA 0.96	1.0000**	0.1027*	1.0000**	1.0000**
Normal EWMA 0.99	0.0013	0.0204	0.0003	0.0006
t-EWMA 0.99	0.0021	0.0345	0.0003	0.0006
Normal Constant 30	0.0046	0.0030	0.0005	0.0006
t-Constant 30	0.0053	0.0079	0.0005	0.0010
Normal Constant 90	0.0230	0.0016	0.0017	0.0021
t-Constant 90	0.0063	0.0021	0.0005	0.0047
Normal GARCH 1	0.0053	0.4201**	0.0000	0.0010
t-GARCH 1	0.0027	0.4201**	0.0000	0.0021
Normal EGARCH 1	0.2124*	0.0593	0.0003	0.1187*
t-EGARCH 1	0.0046	0.0607	0.0001	0.0135
Normal GJR 1	0.0053	0.3430**	0.0001	0.0010
t-GJR 1	0.0013	0.3027**	0.0001	0.0010
Normal EWMA 0.94 1	0.0771	0.3078**	0.0001	0.0030
t-EWMA 0.94 1	0.0170	0.3430**	0.0001	0.0036
Normal GARCH 2	0.0046	0.1322*	0.0001	0.0021
Normal EGARCH 2	0.2251*	0.0590	0.0005	0.2432*
Normal GJR 2	0.0046	0.1322*	0.0001	0.0021
Normal EWMA 0.94 2	0.1153*	0.1322*	0.0001	0.0535

the CME margin value is 100.[23]

We again employ the WTI term structure data and in addition we construct synthetic, 3-month maturity futures on three other contracts: the S&P500 index, gold, and the EUR/USD exchange rate. Table 5 reports summary statistics for these data. All exhibit high volatility around the time of the banking crisis between September 2008 and March 2009, when contagion spread across most financial markets. WTI is far more volatile than other futures of similar maturities, the least volatile being the Euro/Dollar currency futures. However, the Euro/Dollar futures have been more volatile since the

[23] The margin for gold futures is \$4,200 and the contract size is 100 Troy ounces; for EUR/USD the margin is \$2,750 is and the contract size is 125,000 Euros; for WTI the margin is currently between \$2,300 and \$2,700 depending on maturity and the contract size is 1,000 barrels.

Table 5: **Summary Statistics for Daily Returns on Synthetic 3-month Maturity Futures**
Summary statistics for daily returns on synthetic 3-month maturity futures WTI Crude Oil (col. 2); Gold (col. 3); Euro (col. 4); and S&P 500 (col. 5). In percentage points.

	WTI 3M	GC 3M	EC 3M	SP 3M
Average daily return	-0.01	0.03	-0.01	0.06
Standard deviation	2.04	1.15	0.63	1.11
Skewness	0.03	-0.38	-0.02	-0.19
Kurtosis	5.66	8.30	4.33	7.37
Min return	-9.95	-9.35	-2.54	-7.24
Max return	10.15	7.82	3.20	6.97

sovereign debt crisis in Europe in 2011 and the Brexit vote in 2016. Interestingly, the WTI 3-month futures exhibit a small positive skew, all other series have negative skewness. The S&P 500 and gold have the greatest kurtosis.

For each underlying, and for each of the MTL series (i) and (ii) above, we now investigate how different calibration objectives influence the optimal values for the parameters. We employ a sample period from 2 January 2009 to 7 August 2016 to calibrate the parameters a, b, c of (6), using the three different objectives described in Section 4. The corresponding margins are labeled E, T and S respectively, in the results. RMSE (E) minimizes the root-mean-square-error between our margin and the CME margin; Target Statistics (T) minimizes the sum of squared normalized deviations of margin properties from some (subjectively chosen) targets. Here we choose targets that mimic the CME margin, viz. the average CME margin, the number of CME margin changes, the average CME margin increase and the average CME margin decrease. We choose a standard deviation of 2 for the number of margin changes, the average margin level standard deviation is set to 0.1 and we use the empirical standard deviations of the CME margin decreases and increases for the normalization of the remaining two statistics. The Stability objective (S) simply minimizes the number of margin changes.

To reduce the very large number of results to a paper of limited length, the tolerance parameters τ_{low} and τ_{high} are not calibrated; instead we fix $\tau_{_\text{low}} = \tau_{_\text{high}} = \tau = 10$ or 20 days. This is a subjective choice, made simply because it gave realistic results. More generally, the tolerance parameters would depend on the CCP's risk attitude, with lower values leading to more frequent margin re-sets. In a non-competitive environment the exchange should be able to make margin changes more risk sensitive without losing trades, hence selecting a low tolerance setting; but when exchanges compete for volume they may choose to limit margin increases with a higher tolerance setting, until such time as they are impelled to raise margins during a period of sustained high volatility. The operational cost to the CCP for implementing margin decreases may also be a factor to consider when setting $\tau_{_\text{high}}$.

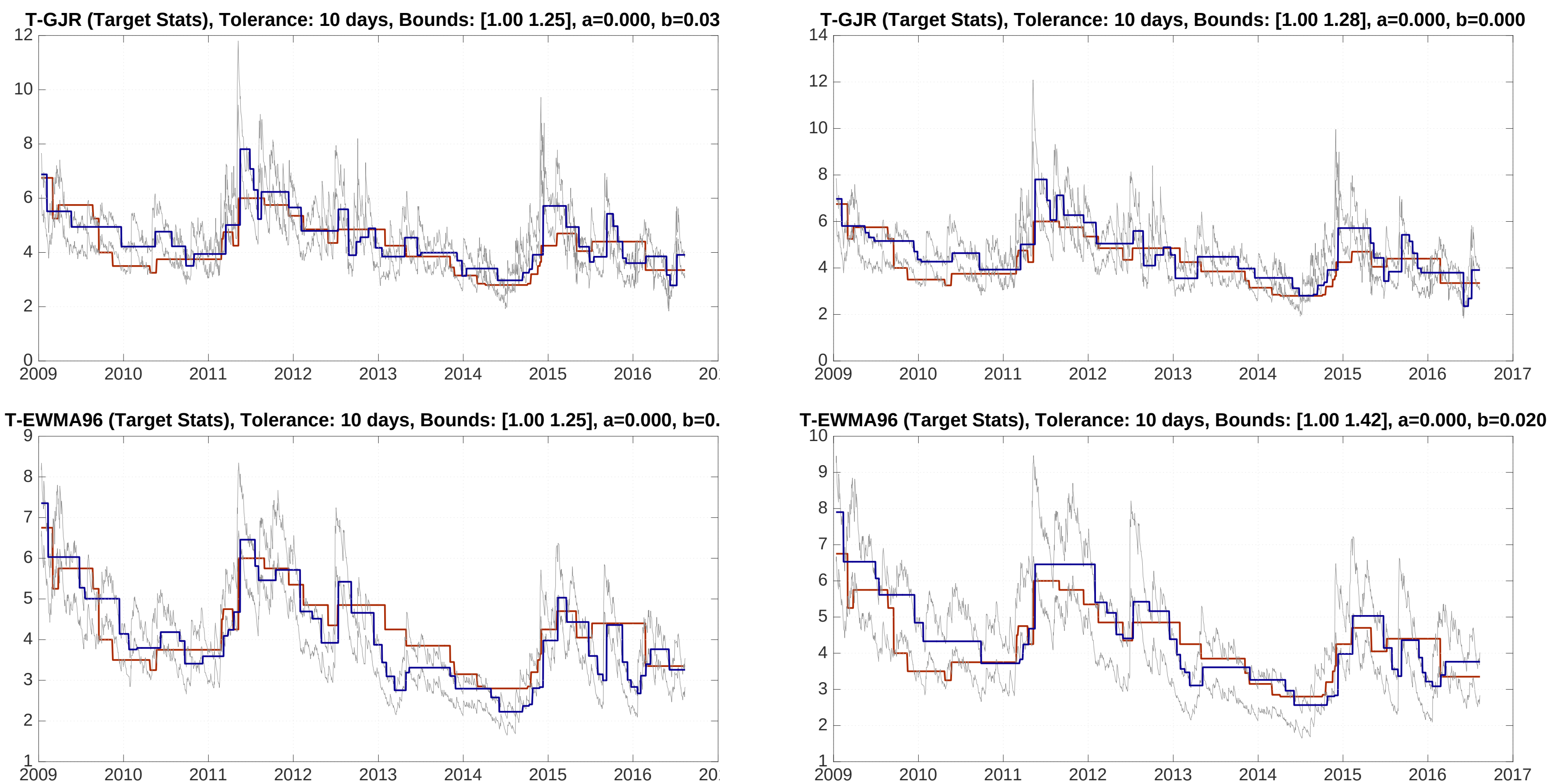
To investigate how sensitive the margin is to the choice of MTL model, for each objective we calibrate the margin model parameters a, b, c twice, once using the MTL forecasts generated by the t-GJR model and again using the MTL from the t-EWMA with $\lambda = 0.96$. Our purpose here is to ask how important it is to use the best possible risk model for margin setting. Will a simple model (t-EWMA) perform as well as a complex model (t-GJR)?

Having calibrated the parameters from all three objectives we now compare our results with CME margins, graphically and statistically. With different calibration objectives, MTL models, and tolerance parameters there are numerous possibilities to consider for each margin, and again it is not possible to include all comparisons on all futures contracts, but some are illustrated below.

Figure 2 depicts results from calibrating the margin model to the 3-month WTI futures series using the T calibration objective. In each figure we show daily time series of upper and lower boundaries, MTL and $(1 + c)$MTL, and state the values of other model parameters above each chart. They allow a visual comparison of margins produced using the same margin model and calibration objective but different MTL models. To economize on space, we report detailed results for the E and S calibration in the Appendix (Figures A.1 and A.2).

Figure 2: **Comparison of Margins with Target Stats Calibration Objective: 3-month WTI Futures**

A comparison of our margin (blue) with the CME margin (red) for 3-month WTI futures. Each figure depicts our margin and the upper and lower bounds MTL and $(1+c)$MTL. The MTL model is either the Student t-GJR, with parameters calibrated to the underlying futures returns (upper) or Student t-EWMA with $\lambda = 0.96$ (lower). The calibrated values for the proportional margin decrease a and increase b are reported above each chart. The calibration objective here is Target Statistics, as explained in the text.



Similar graphs for the term structure of Euro/Dollar futures and of gold futures (not presented here but available on request) confirm that our margins tend to be higher and less stable when the tolerance is set at 10 days, and when we employ the t-EWMA model for MTL rather than the t-GJR. The CME S&P 500 margins are extremely high and seem to bear no relation to risk whatsoever, as previously documented by Abruzzo and Park (2016).[24]

Table 6 presents some statistics that provide a more complete comparison between the CME and our margins for WTI futures. Again we need to restrict the results, for lack of space, and so we only present results for long positions, i.e. using the left-hand tails of the returns distribution. We also report only one tolerance parameter setting, i.e. 20 days.[25] The results are disaggregated according to the futures contract, the MTL model, the calibration objective and the different settings for the tolerance. In column 1 the margin is denoted first by the objective, then the tolerance setting and then by either E or G according as the t-EWMA or the t-GJR model is used to forecast the MTL. The next three columns report the calibrated values of the margin model parameters and subsequent columns provide the summary statistics for the number of changes (divided into increases and decreases), the average number of days between changes, the average % increase and % decrease, the smallest increase and the smallest decrease.

In these tables, the first six rows report results based on our margin models, labelled {SE, SG, TE, TG, EE, EG}. The first letter denotes the calibration objective (S, T or E as described above) and the second letter denotes the risk model (E for the t-EWMA and G for t-GJR GARCH). The next six rows present the results of calibrating the margin scheme (5) of Lam *et al.* (2010) using the same objectives, and with both t-EWMA and t-GJR models for estimating the volatility parameters. These are denoted *LE and *LG respectively, where * denotes the calibration objective (S, T or E).[26] The penultimate two rows give the results for the Chiu *et al.* (2006) margin scheme (4), with VaR being computed using the t-EWMA and t-GJR models, and these are denoted CE and CG respectively. Finally, in bold, in the bottom row we report the statistics corresponding to the CME margins. In each case we also mark in bold the name of the model(s) that exhibit statistical features which are closest to those of the CME margins.

It is clear that the WTI margins produced using the rule of Chiu *et al.* (2006) are the most unstable. This is also true for other contracts for which we provide detailed results in the Appendix (Tables A.3 to A.6). They change much more often than the other margins, especially the CME margins. By contrast, the margins based on the scheme of Lam *et al.* (2010) are sometimes too stable, when calibrated using the stability criteria. They also exhibit much larger increases then other margins, and

[24]Abruzzo and Park (2016) find that the probability of margin shortfall is only 0.03 for the S&P 500, compared with 0.13 for gold and 0.66 for WTI. Perhaps there is insufficient competition in this market to drive down margins. But holding S&P 500 margins so high could also be a policy decision to reduce excessive speculation on an index which dominates the health of the US economy.

[25]Results for lower levels of tolerance are available on request. In every case, the higher tolerance value $\tau_{\text{low}} = \tau_{\text{high}} = 20$ matches the number of CME margin changes better than the lower values. Clearly, the CME does not change margins immediately following a change in volatility.

[26]These are not the volatility models and calibration objectives that were employed by Lam *et al.* (2010). We adopt our volatility models and objectives here so that the focus of the comparison is on the margin scheme itself and not on other aspects of the model.

Table 6: **Summary Statistics on Margins for 1-month and 3-month WTI Futures**

This table reports statistics for 1-month and 3-month WTI futures margins based on a tolerance setting $\tau_{\text{low}} = \tau_{\text{high}} = 20$ days and for the EWMA and *t*-GJR MTL models. The model name in column 1 starts with the margin objective (S, T or E) followed by E or G, according to the MTL model (either EWMA or *t*-GJR GARCH). The same objectives are used for the scheme introduced by Lam *et al.* (2010), and these models have an additional L (e.g. SLE denotes the Lam *et al.* (2010) model with stability objective based on EWMA). The other margin schemes are of Chiu *et al.* (2006), also with two possible MTL models (denoted CE and CG) and, finally the CME. Columns 2 – 4 report the calibrated parameters and subsequent columns provide summary statistics.

	a	b	c	No. of changes	No. of increases	No. of decreases	Avg. days between changes	Avg. margin increase (%)	Avg. margin decrease (%)	Smallest increase (%)	Smallest decrease (%)
Panel A: WTI (1M)											
SE	0.03	0.16	0.25	21	10	11	92.10	26.75	-24.49	6.51	-13.89
SG	0.00	0.02	0.25	22	8	14	88.90	17.33	-12.75	1.52	-2.51
TE	0.00	0.00	0.31	30	10	20	63.66	18.76	-10.88	3.06	-0.06
TG	0.00	0.00	0.25	20	6	14	97.00	19.25	-11.38	3.07	-0.41
EE	0.01	0.01	0.25	27	10	17	70.92	21.66	-14.17	3.06	-7.01
EG	**0.00**	**0.02**	**0.25**	**23**	**9**	**14**	**84.86**	**15.15**	**-12.67**	**1.40**	**-2.51**
SLE	2.99	0.46	–	6	3	3	286.40	52.62	-46.54	46.33	-46.15
SLG	2.90	0.49	–	3	1	2	293.50	127.03	-49.34	127.03	-49.22
TLE	2.57	0.25	–	27	12	15	67.54	37.96	-25.67	25.49	-25.05
TLG	2.43	0.25	–	38	19	19	50.43	32.42	-26.51	25.58	-25.07
ELE	2.56	0.25	–	27	12	15	67.54	37.96	-25.67	25.49	-25.05
ELG	2.30	0.25	–	34	17	17	56.55	33.14	-27.15	25.58	-25.59
CE	–	–	–	70	30	40	27.29	22.96	-15.93	15.02	-15.03
CG	–	–	–	105	47	58	18.25	24.11	-16.77	15.01	-15.08
CME			**–**	**30**	**12**	**18**	**60.66**	**13.53**	**-11.00**	**1.72**	**-4.00**
Panel B: WTI (3M)											
SE	0.02	0.18	0.25	22	11	11	83.86	26.86	-25.69	5.61	-16.15
SG	0.00	0.03	0.25	20	8	12	90.05	20.27	-15.65	2.28	-5.50
TE	0.00	0.00	0.33	27	9	18	68.19	19.40	-11.71	3.68	-3.33
TG	**0.00**	**0.00**	**0.25**	**20**	**6**	**14**	**97.11**	**23.63**	**-12.42**	**4.50**	**-2.03**
EE	0.00	0.04	0.26	26	9	17	70.36	24.61	-14.51	3.22	-4.40
EG	**0.00**	**0.01**	**0.25**	**20**	**7**	**13**	**97.11**	**21.55**	**-13.86**	**5.05**	**-2.03**
SLE	2.45	0.49	–	8	4	4	221.29	67.53	-49.08	53.07	-48.89
SLG	2.49	0.45	–	20	11	9	86.26	60.15	-46.07	45.52	-44.85
TLE	2.46	0.25	–	29	13	16	66.39	36.85	-25.55	25.40	-25.06
TLG	2.49	0.25	–	48	24	24	40.04	33.50	-27.16	25.31	-25.31
ELE	2.47	0.25	–	29	13	16	66.39	36.85	-25.55	25.40	-25.06
ELG	2.36	0.27	–	43	22	21	44.81	33.96	-28.79	26.91	-26.53
CE	–	–	–	74	33	41	25.82	21.28	-15.94	15.01	-15.02
CG	–	–	–	152	69	83	12.60	23.85	-16.77	15.01	-15.01
CME			**–**	**31**	**13**	**18**	**58.63**	**12.81**	**-11.42**	**1.79**	**-1.75**

this would act as a deterrent to investors. In each table the margins that are closest to the CME's are ours, based on a tolerance of 20 days, usually the T (sometimes the E) calibration objective, and usually the t-GJR (sometimes the t-EWMA) MTL model.

For our margin schemes the number of changes decreases and their frequency increases as the tolerance parameter increases, a feature that is built into the design of our margins. Because of this and the intrinsic features of MTL, the average number of increases is greater than the average number of decreases, while the average size of increases is smaller than the average size of decreases. These features are shared by the CME margins, as previously noted by Abruzzo and Park (2016) and Daskalaki and Skiadopoulos (2016).

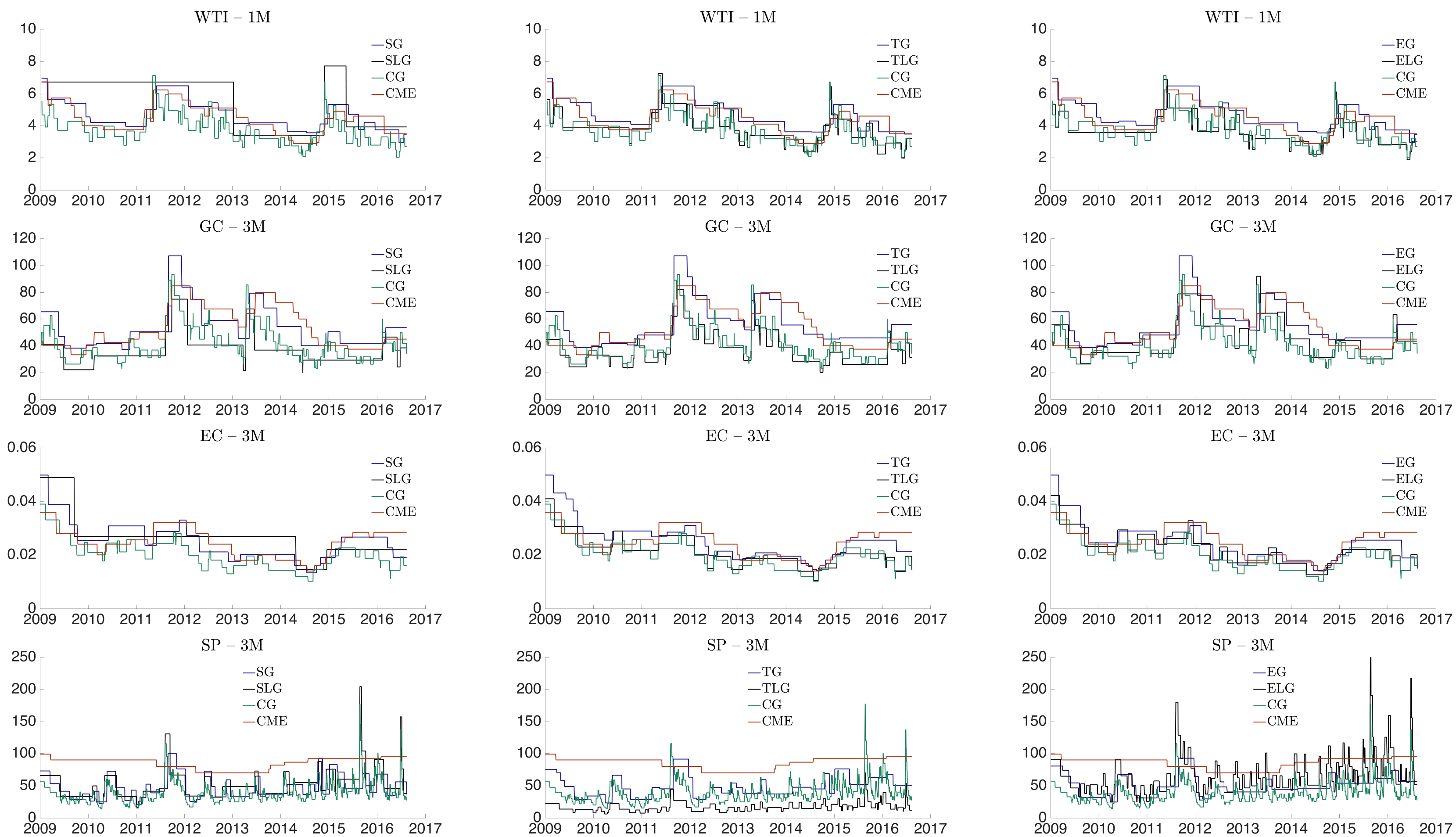
Notably, changes in the CME's S&P 500 futures margin occur much less often than in other markets, as already noted by Abruzzo and Park (2016) and Daskalaki and Skiadopoulos (2016), with only 5 increases and 3 decreases over our entire sample and an average of 247.29 days between changes. WTI and Euro-dollar currency futures margins change more often, but even for these futures there are only about 4 margin changes per year. When CME margins change they increase/decrease by about 10-13%, depending on the contract; the exceptions are increases in gold margins which average 17.78% and S&P margin increases, which are only 6.31% on average. All CME margins (except those for the S&P 500 futures) decrease more often than they increase. Hence their common pattern is one of slowly stepping decreases and larger but less frequent increases.

Figures 3 provides a visual comparison of the margins for the prompt futures on each underlying, based on the t-GJR volatility model and using all three calibration objectives. There are several noteworthy observations. First, the scheme proposed by Lam *et al.* (2010) appears to be incompatible with the stability calibration criteria. That is, the margins corresponding to SLYWG have too few changes, and consequently remain for a very long time either far above the CME margins (e.g. for 1M WTI and for 3M EC) or far below the CME margins (for 3M WTI). Secondly, both the margins proposed by Lam *et al.* (2010) and those of Chiu *et al.* (2006) only remain unchanged when the quantile remains within a very strict corridor. Consequently, they tend to produce an excessive number of margin changes and, more importantly, changes are often reversed after only a few days. By contrast, our margin schemes are designed to stabilize margins and to prevent frequent margin reversals. The CME margins are positively autocorrelated, i.e. a margin increase is more likely to be followed by a further increase than a decrease (and vice versa). This is an important feature which is also built into our margin schemes.

Thirdly, our average margin levels remain very close to CME margins for all but the S&P 500 futures, whereas the rules of Chiu *et al.* (2006) and Lam *et al.* (2010) typically produce much lower margins than the CME. As previously discussed, the CME margins for S&P 500 futures appear to be unrelated to risk, but our scheme proposes margins that have very similar characteristics to the CME margins for WTI, gold and Euro/Dollar futures. This suggests that our proposed scheme would provide risk-sensitive, yet stable margin requirements for other products. Our schemes also work well across different futures contracts and futures maturities, which also suggests that they are adaptable to the characteristics of diverse underlying contracts and market conditions.

Figure 3: **Comparison of Margins.**

A comparison of our margin (in blue) with the CME margin (red) and the margins obtained using Chiu *et al.* (2006) (green) and Lam *et al.* (2010) (black) for 1-month WTI futures, 3-month Gold Futures, 3-month Euro/Dollar Futures and 3-month S&P 500 Futures using Student t-GJR volatility model. The calibration objectives are Stability (left), Target Statistics (middle) and RMSE (right).



5.3 Further Results and Discussion

In this section, we provide a brief summary of additional robustness tests.[27] First, we ask whether the calibrated margin models (rather than the MTL models) pass our back-testing procedure. One difficulty with coverage tests of the margin models in our framework is that margin models use an MTL estimate to determine only one of the bounds of our margin corridor. This implies that the margin model is by construction much more conservative since our margins are forced to stay above the absolute MTL level. Margins may only cross this threshold for brief periods of time, as measured by our tolerance parameter τ. This asymmetry implies that our margin models will probably fail coverage tests – not because there are too many exceptions but because they are, by construction, expected to produce *fewer* exceptions – at the MTL coverage level. Un-tabulated empirical results confirm our margin models are more conservative than those of Chiu *et al.* (2006) and Lam *et al.* (2010). With some exceptions, their margins are not rejected by coverage tests – but this is at our chosen MTL coverage levels. In fact, because of the rules we choose to construct them, the true coverage level for our margins should be *smaller* than the coverage level set for the MTL back-tests. However, it is not clear which coverage level is adequate for a formal test of the margin model.

As a second robustness check, we consider the use of at-the-money (ATM) implied volatility in the design of our margin models. We do not incorporate this analysis into our main results as implied volatility data are much more scarce: for WTI and gold, ATM implied volatilities are only available (on Bloomberg) from October 2005, whereas for Euro-USDollar currency futures the ATM implied volatility data are available from December 1998.[28] These limitations restrict our back-testing sample considerably, compared with our earlier study on WTI, and hence we only outline these results.

We estimate a parsimonious time-series model which incorporates at-the-money implied volatility (we obtain implied volatility data from Bloomberg) by setting $\sigma^2_{t,T} = \psi \times \theta^2_{t-1,T}$, where $\theta_{t,T}$ is the implied volatility at time t for an ATM option with time to maturity T, and ψ is a constant model parameter. It is well-known from the literature on option risk premia (see Carr and Wu (2008)) that option prices contain volatility and jump risk premia which leads to an upward bias in implied volatility. The parameter ψ allows us to capture this effect and we expect $\psi < 1$.

The MTL is obtained using equation (1), under the assumption that the conditional distribution of returns is normal, or using (2), when we assume that their conditional distribution is Student t. This way, the implied volatility models are based on only one or two parameters: ψ and, for models with Student t-distributed error terms, the degrees of freedom parameter. We calibrate these parameters using a (rolling) window of 750 days, rather than 1500 days (as required for the GARCH models). This choice is helpful, because implied volatility data are not available over our entire sample, which is exceptionally long. Futures price data are available, but the implied volatility time series obtainable from Bloomberg start several years later.[29]

[27] We would like to thank an anonymous referee for suggesting these exercises.

[28] Due to the lack of risk-sensitivity of the S&P 500 futures margins, we drop these data from the analysis.

[29] We have also experimented with other possible model specifications, and on combinations of implied volatility with standard GARCH specifications in particular. However, the calibration of these models requires a longer rolling window to obtain stable results. Therefore, given the comment just made about the availability of implied volatility time series,

We find that the parsimonious implied volatility models introduced above perform well in the Christoffersen (1998) two-sided interval tests for MTL. These findings are in line with a large literature that finds incremental information in option prices which are based on forward-looking information (see Jiang and Tian (2005)). However, a comparison between margin models based on implied volatility MTL, versus GARCH or EWMA MTL, shows only marginal differences between the margins obtained. Examples are included in the online appendix, Table A.7.

6 Conclusions

We extend the literature on new margin schemes for derivatives in several important respects. Our first innovation is to employ MTL as the risk metric instead of VaR or ETL, because MTL is less sensitive to outliers than ETL, and because it is relatively straightforward to back-test. This electability is a major advantage of MTL over ETL. In fact, it has recently been proved that ETL is jointly elicitable with VaR (Ziegel, 2016) but an effective simple back-testing procedure has yet to be developed.

Then we back-test both tails simultaneously (because the CCP will have both long and short exposures to default) implementing the state-of-the art testing procedures introduced by Gneiting (2011) and Hansen *et al.* (2011). Many conditional parametric models with time-varying volatility lie in the same model confidence set. Only those with constant (or almost constant) volatility perform poorly. Based on the conditional ranked probability score, asymmetric GARCH models with t-distributed innovations are superior to other MTL models. Yet in the model confidence sets there is scant evidence of superiority over the t-EWMA family, with an appropriate choice of smoothing constant. For simplicity, CCPs may prefer to implement a EWMA model such as the t distributed MTL with EWMA volatility based on a smoothing constant of 0.96. However, in almost all cases our margins match the CME margins better when we employ an MTL derived from a Student-t asymmetric GARCH model.

We have introduced a simple parametric rule for setting margins along a term structure of futures contracts which requires a maximum of five parameters. These parameters govern the boundaries for a margin reset, the time period before a reset, and it's size when a reset is triggered. They may be calibrated to a time-series of MTL and – if required – an historical time-series for a target (e.g. the CME margin) using a variety of objectives. Our empirical study has compared our margins with two other schemes, advocated by Chiu *et al.* (2006) and Lam *et al.* (2010) respectively. Those schemes tend to produce unstable margins which are generally lower than CME margins. By contrast, we are even able to mimic most of the characteristics of CME margins with just three calibrated parameters. Moreover, by comparing our model for different assets with SPAN margins we add useful information, e.g. that SPAN margins for S&P 500 futures are too risk-insensitive, although our approach can produce margins that are very close to SPAN in several other markets such as gold and the Euro-US Dollar exchange rate.

this reduces the amount of data for back-testing quite significantly. Moreover, based on the data which are available for back-testing, there is no significant improvement in the margin that are obtained using a more sophisticated model. Hence, we focus on the simple model specification introduced above and we find that, in line with the literature on volatility forecasting, implied volatility models perform well in backtests, overall.

References

Abruzzo, N. and Park, Y.-H. (2016). An empirical analysis of futures margin changes: Determinants and policy implications. *Journal of Financial Services Research*, **49**, 65–100.

Adrian, T. and Shin, H. S. (2014). Procyclical leverage and value-at-risk. *Review of Financial Studies*, **27**(2), 373–403.

Alexander, C. (2001). Orthogonal GARCH. *Mastering Risk*, **2**, 21–38. Financial Times-Prentice Hall, London.

Alexander, C. (2008). Value-at-risk models. *Market Risk Analysis, Volume IV*. Wileys.

Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (2002). Backtesting derivative portfolios with filtered historical simulation. *European Financial Management*, **8**(1), 31–58.

BIS (2015). Margin requirements for non-centrally cleared derivatives. *Available at http://www.bis.org/bcbs/publ/d317.pdf*.

Booth, G. G., Broussard, J. P., Martikainen, T., and Puttonen, V. (1997). Prudent margin levels in the Finnish stock index futures market. *Management Science*, **43**, 1177–1188.

Brennan, M. J. (1986). A theory of price limits in futures markets. *Journal of Financial Economics*, **16**(2), 213–233.

Broussard, J. P. and Booth, G. (1998). The behavior of extreme values in Germany's stock index futures: An application to intradaily margin setting. *European Journal of Operational Research*, **104**(3), 393–402.

Brunnermeier, M. K. and Pedersen, L. H. (2009). Market liquidity and funding liquidity. *Review of Financial Studies*, **22**(6), 2201–2238.

Carr, P. and Wu, L. (2008). Variance risk premiums. *The Review of Financial Studies*, **22**(3), 1311–1341.

Chen, Q., Gerlach, R., and Lu, Z. (2012). Bayesian value-at-risk and expected shortfall forecasting via the asymmetric Laplace distribution. *Computational Statistics & Data Analysis*, **56**(11), 3498 – 3516.

Chiu, C.-L., Chiang, S.-M., Hung, J.-C., and Chen, Y.-L. (2006). Clearing margin system in the futures markets: Applying the value-at-risk model to Taiwanese data. *Physica A: Statistical Mechanics and its Applications*, **367**(0), 353 –374.

Chou, R., Wang, G. H., and Wang, Y.-Y. (2015). The effects of margin changes on the composition of traders and market liquidity: Evidence from the Taiwan futures exchange. *Journal of Futures Markets*, **50**, 894–915.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, **39**(4), 841 – 862.

Clements, M., Galvão, A., and Kim, J. (2008). Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance*, **15**(4), 729–750.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, **7**(2), 174–196.

Cotter, J. (2001). Margin exceedences for European stock index futures using extreme value theory. *Journal of Banking and Finance*, **25**(8), 1475 – 1502.

Daníelsson, J., Jorgensen, B. N., Samorodnitsky, G., Sarma, M., and de Vries, C. G. (2013). Fat tails, VaR and subadditivity. *Journal of Econometrics*, **172**(2), 283–291.

Daskalaki, C. and Skiadopoulos, G. (2016). The effect of margin changes on commodity futures. *Journal of Financial Services Research*, **22**, 129–152.

Day, T. E. and Lewis, C. M. (1997). Initial margin policy and stochastic volatility in the crude oil futures market. *Review of Financial Studies*, **10**(2), 303–332.

Duffie, D. and Zhu, H. (2011). Does a central clearing counterparty reduce counterparty risk? *Review of Asset Pricing Studies*, **1**(1), 74–95.

Eckardt, Walter L., J. and Rogoff, D. L. (1976). 100% margins revisited. *Journal of Finance*, **31**(3), 995–1000.

Embrechts, P., Resnick, S. I., and Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, **3**(2), 30–41.

European Union (2013). Commission delegated regulation (EU) no 153/2013. *Official Journal of the European Union*.

Figlewski, S. (1984). Margins and market integrity: Margin setting for stock index futures and options. *Journal of Futures Markets*, **4**(3), 385–416.

Galai, N. (1979). A proposal for indexes for traded call options. *Journal of Finance*, **34**, 1157–1172.

Garleanu, N. and Pedersen, L. H. (2011). Margin-based asset pricing and deviations from the law of one price. *Review of Financial Studies*, **24**(6), 1980–2022.

Gay, G. D., Hunter, W. C., and Kolb, R. W. (1986). A comparative analysis of futures contract margins. *Journal of Futures Markets*, **6**(2), 307–324.

Gibson, R. and Murawski, C. (2013). Margining in derivatives markets and the stability of the banking sector. *Journal of Banking and Finance*, **37**, 1119–1132.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, **48**(5), 1779–1801.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, **106**(494), 746–762.

Gneiting, T. and Ranjan, R. (2011). Comparing Density Forecasts Using Threshold- and Quantile-Weighted Scoring Rules. *Journal of Business and Economic Statistics*, **29**(3), 411–422.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, **79**(2), 453–497.

Hardouvelis, G. A. (1990). Margin requirements, volatility, and the transitory component of stock prices. *American Economic Review*, **80**(4), 736–762.

Hardouvelis, G. A. and Theodossiou, P. (2002). The asymmetric relation between initial margin requirements and stock market volatility across bull and bear markets. *Review of Financial Studies*, **15**(5), 1525–1559.

Heckinger, R., Cox, R., and Marshall, D. (2016). Cleared margin setting at selected CCPs. *Federal Reserve Bank of Chicago, Economic Perspectives*, **4**.

Hsieh, D. A. and Miller, M. H. (1990). Margin regulation and stock market volatility. *Journal of Finance*, **45**(1), 3–29.

Jiang, G. J. and Tian, Y. S. (2005). The model-free implied volatility and its information content. *The Review of Financial Studies*, **18**(4), 1305–1342.

Kou, S. and Peng, X. (2014a). Expected shortfall or median shortfall. *Journal of Financial Engineering*, **1**(01).

Kou, S. and Peng, X. (2014b). On the measurement of economic tail risk. *Available at https://arxiv.org/pdf/1401.4787.pdf*.

Kou, S., Peng, X., and Heyde, C. C. (2013). External risk measures and Basel accords. *Mathematics of Operations Research*, **38**(3), 393–417.

Kumar, R., Ferris, S. P., and Chance, D. M. (1991). The differential impact of Federal Reserve margin requirements on stock return volatility. *Financial Review*, **26**(3), 343–366.

Kupiec, P. H. (1993). Futures margins and stock price volatility: Is there any link? *Journal of Futures Markets*, **13**(6), 677–691.

Kupiec, P. H. (1994). The performance of S&P 500 futures product margins under the SPAN margining system. *Journal of Futures Markets*, **14**(7), 789–811.

Kupiec, P. H. and White, A. P. (1996). Regulatory competition and the efficiency of alternative derivative product margining systems. *Journal of Futures Markets*, **16**(8), 943–968.

Lam, K., Sin, C.-Y., and Leung, R. (2004). A theoretical framework to evaluate different margin-setting methodologies. *Journal of Futures Markets*, **24**(2), 117–145.

Lam, K., Yu, P., and Lee, P. (2010). A margin scheme that advises on when to change required margin. *European Journal of Operational Research*, **207**(1), 524–530.

Largay, J. A. and West, R. R. (1973). Margin changes and stock price behavior. *Journal of Political Economy*, **81**(2), 328–339.

Longin, F. M. (1999). Optimal margin level in futures markets: Extreme price movements. *Journal of Futures Markets*, **19**(2), 127–152.

Louzis, D., Xanthopoulos-Sisinis, S., and Refenes, A. (2014). Realized volatility models and alternative value-at-risk prediction strategies. *Economic Modelling*, **40**, 101–116.

Mägerle, J. and Nellen, T. (2011). Interoperability between central counterparties. *Swiss National Bank*. Working Paper 2011-12.

Murphy, D., Vasios, M., and Vause, N. (2014). An investigation into the procyclicality of risk-based initial margin models. Bank of England Financial Stability Paper No. 29.

Murphy, D., Vasios, M., and Vause, N. (2016). A comparative analysis of tools to limit the procyclicality of initial margin requirements. Bank of England Staff Working Paper No. 597.

Nadarajah, S., Zhang, B., and Chan, S. (2014). Estimation methods for expected shortfall. *Quantitative Finance*, **14**(2), 271–291.

Nahai-Williamson, P., Ota, T., Vital, M., and Wetherilt, A. (2013). Central counterparties and their financial resources - a numerical approach. Bank of England Financial Stability Paper No. 19.

Nelson, D. B. (1990). ARCH models as diffusion approximations. *Journal of Econometrics*, **45**(1-2), 7 – 38.

Nieto, M. and Ruiz, E. (2016). Frontiers in var forecasting and backtesting. *International Journal of Forecasting*, **32**(2), 475–501.

Phylaktis, K. and Aristidou, A. (2013). Margin changes and futures trading activity: a new approach. *European Financial Management*, **19**, 45–71.

Rytchkov, O. (2016). Time-varying margin requirements and optimal portfolio choice. *Journal of Financial and Quantitative Analysis*, **51**(2), 655–683.

Sèvi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research*, **235**, 643–659.

Su, Y. C., Huang, H. C., and Lin, Y. J. (2011). GJR-GARCH model in value-at-risk of financial holdings. *Applied Financial Economics*, **21**(24), 1819–1829.

Telser, L. G. (1981). Margins and futures contracts. *Journal of Futures Markets*, **1**(2), 225–253.

Ziegel, J. F. (2016). Coherence and elicitability. *Mathematical Finance*, **26**(4), 901–918.

A Online Appendix

Table A.1 shows results of Gneiting and Ranjan (2011) tests for pairwise comparison of the forecasting performance of all models. These tests are based on the t_{ij}-statistic described in the main paper (Section 3.2). We use the weight function $w(\alpha) = (2\alpha - 1)^2$ because our focus is on accuracy in both tails simultaneously. Here the model in row i is preferred to the model in column j if the corresponding statistic is negative, and the lower and more negative the statistic the greater the accuracy of model i relative to model j. Clearly, the best performing model is the GJR with Student t innovations, which is labeled model (6) in the table, but the t-EGARCH model (4) performs almost as well. The worst is model (12), the t-EWMA with smoothing constant 0.99 but model (14), the student t distribution with standard deviation based on a sample of size 30, is almost as poor. Indeed, all models (11) – (16) are outperformed by most of the others.

Table A.2 completes this robustness check for the multivariate term structure models, labeled (17) – (28) in Table A.1. Again we confirm the conditional coverage test results in Table 2: orthogonal models with only one component, t-distributed innovations and a GARCH volatility structure predict both tails simultaneously with sufficient accuracy, and all models based on two principal components are inferior.

There is no space to present results for other weight functions and other maturities of WTI futures (these are available on request) so here we just summarize their conclusions. First, with $w(\alpha) = 1$, all the constant-volatility models perform particularly badly, and all models with normal innovations are outperformed by those with Student-t innovations. The GARCH models with t-distributed error terms significantly outperform most of the other specifications. The best performing models with $w(\alpha) = 1$ also perform strongly in the left tail, with $w(\alpha) = (1 - \alpha)^2$, but the EWMA models with Student-t innovations and smoothing constants of 0.94 or 0.96 also perform reasonably well. For right-tail forecasting, i.e. with $w(\alpha) = \alpha^2$, the normal EGARCH model is better than the t-EGARCH or t-GJR models, which themselves out-perform all models except normal EGARCH. This finding is consistent with a distribution having asymmetric tails. Indeed, in most financial markets the left tail typically has greater probability mass than the right tail.

The remaining figures and tables in this appendix complement the main body of the paper with additional results and robustness checks.

Table A.1: **Gneiting-Ranjan Tests with weighting in both tails.**

This table reports the Gneiting and Ranjan (2011) statistic t_n for comparing all MTL model pairs for 1-month WTI futures, with $w(\alpha) = (2\alpha - 1)^2$. The test data sample is from 1 Jan 1996 to 31 Dec 2008. A positive statistic for a comparison of model x (in row) with model y (in column) indicates that the model in the row is out-performed by the model in the column and vice versa. The asymptotic distribution of the t_n-statistic is standard normal.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Normal GARCH (1)	–	2.00	0.17	1.86	0.40	2.14	0.20	0.20	1.55	1.60	-1.37	-1.19	-2.11	-1.99	-0.94	-0.91
t-GARCH (2)	-2.00	–	-1.51	0.17	-1.53	0.70	-1.40	-1.40	-0.20	-0.14	-2.98	-2.83	-3.77	-3.68	-2.72	-2.71
Normal EGARCH (3)	-0.17	1.51	–	2.10	0.06	1.95	0.06	0.06	1.18	1.21	-1.46	-1.28	-2.04	-1.94	-1.02	-0.98
t-EGARCH (4)	-1.86	-0.17	-2.10	–	-1.65	0.38	-1.29	-1.29	-0.25	-0.21	-3.00	-2.85	-3.42	-3.34	-2.58	-2.58
Normal GJR (5)	-0.40	1.53	-0.06	1.65	–	1.94	0.02	0.03	1.27	1.31	-1.40	-1.23	-2.15	-2.04	-1.02	-0.98
t-GJR (6)	-2.14	-0.70	-1.95	-0.38	-1.94	–	-1.63	-1.63	-0.51	-0.46	-3.00	-2.85	-3.85	-3.76	-2.74	-2.72
Normal EWMA 0.94 (7)	-0.20	1.40	-0.06	1.29	-0.02	1.63	–	3.21	3.00	3.10	-1.18	-1.05	-3.94	-3.73	-1.00	-0.96
t-EWMA 0.94 (8)	-0.20	1.40	-0.06	1.29	-0.03	1.63	-3.21	–	2.99	3.09	-1.18	-1.05	-3.95	-3.74	-1.00	-0.97
Normal EWMA 0.96 (9)	-1.55	0.20	-1.18	0.25	-1.27	0.51	-3.00	-2.99	–	5.04	-2.33	-2.15	-5.41	-5.23	-2.49	-2.42
t-EWMA 0.96 (10)	-1.60	0.14	-1.21	0.21	-1.31	0.46	-3.10	-3.09	-5.04	–	-2.36	-2.19	-5.47	-5.28	-2.53	-2.46
Normal EWMA 0.99 (11)	1.37	2.98	1.46	3.00	1.40	3.00	1.18	1.18	2.33	2.36	–	3.02	-0.57	-0.49	0.81	0.9
t-EWMA 0.99 (12)	1.19	2.83	1.28	2.85	1.23	2.85	1.05	1.05	2.15	2.19	-3.02	–	-0.70	-0.61	0.50	0.58
Normal Constant 30 (13)	2.11	3.77	2.04	3.42	2.15	3.85	3.94	3.95	5.41	5.47	0.57	0.70	–	3.20	1.06	1.09
t-Constant 30 (14)	1.99	3.68	1.94	3.34	2.04	3.76	3.73	3.74	5.23	5.28	0.49	0.61	-3.20	–	0.96	0.99
Normal Constant 90 (15)	0.94	2.72	1.02	2.58	1.02	2.74	1.00	1.00	2.49	2.53	-0.81	-0.50	-1.06	-0.96	–	0.82
t-Constant 90 (16)	0.91	2.71	0.98	2.58	0.98	2.72	0.96	0.97	2.42	2.46	-0.90	-0.58	-1.09	-0.99	-0.82	–
Normal GARCH 1 (17)	1.03	2.48	0.89	2.30	1.06	2.62	0.98	0.98	2.07	2.12	-0.58	-0.44	-1.36	-1.25	-0.20	-0.17
t-GARCH 1 (18)	-0.80	0.91	-0.62	0.82	-0.59	1.16	-0.64	-0.64	0.48	0.52	-1.94	-1.81	-2.93	-2.85	-1.64	-1.63
Normal EGARCH 1 (19)	0.55	2.00	0.72	2.30	0.72	2.33	0.62	0.62	1.62	1.66	-0.78	-0.64	-1.49	-1.40	-0.40	-0.38
t-EGARCH 1 (20)	-0.36	1.34	-0.26	1.62	-0.21	1.74	-0.20	-0.20	0.86	0.90	-1.55	-1.42	-2.40	-2.31	-1.20	-1.18
Normal GJR 1 (21)	0.40	1.79	0.45	1.80	0.58	2.12	0.50	0.50	1.53	1.56	-0.86	-0.72	-1.66	-1.56	-0.51	-0.48
t-GJR 1 (22)	-0.96	0.48	-0.82	0.53	-0.80	0.87	-0.86	-0.85	0.21	0.25	-2.01	-1.89	-3.07	-2.99	-1.73	-1.72
Normal EWMA 0.94 1 (23)	1.01	2.33	0.99	2.17	1.10	2.50	1.79	1.79	2.84	2.88	-0.22	-0.11	-1.26	-1.13	0.10	0.13
t-EWMA 0.94 1 (24)	1.01	2.33	0.99	2.17	1.10	2.50	1.79	1.79	2.83	2.88	-0.23	-0.11	-1.27	-1.13	0.10	0.12
Normal GARCH 2 (25)	0.97	2.24	0.82	2.12	1.02	2.41	0.89	0.89	2.04	2.08	-0.66	-0.52	-1.47	-1.36	-0.30	-0.27
Normal EGARCH 2 (26)	0.24	1.66	0.41	1.93	0.44	1.99	0.36	0.36	1.42	1.46	-1.01	-0.87	-1.76	-1.66	-0.64	-0.61
Normal GJR 2 (27)	0.35	1.65	0.41	1.67	0.55	1.96	0.45	0.45	1.51	1.55	-0.89	-0.76	-1.71	-1.61	-0.56	-0.53
Normal EWMA 0.94 2 (28)	0.71	1.95	0.72	1.84	0.83	2.13	1.38	1.39	2.56	2.61	-0.47	-0.35	-1.71	-1.57	-0.18	-0.15

Table A.2: **Gneiting-Ranjan Tests with weighting in both tails.**

This table reports the Gneiting and Ranjan (2011) statistic t_n for all MTL model pairs, with $w(\alpha) = (2\alpha - 1)^2$. Test data sample 1 Jan 1996 to 31 Dec 2008. A positive statistic for a comparison of model x (in row) with model y (in column) indicates that the model in the row is out-performed by the model in the column and vice versa. The asymptotic distribution of the t_n-statistic is standard normal.

	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)
Normal GARCH 1 (17)	–	2.23	0.50	1.50	1.38	2.39	-0.52	-0.52	0.44	1.02	1.17	-0.01
t-GARCH 1 (18)	-2.23	–	-1.66	-1.03	-1.39	0.86	-2.49	-2.48	-1.72	-1.14	-1.13	-1.76
Normal EGARCH 1 (19)	-0.50	1.66	–	1.40	0.33	1.99	-0.75	-0.75	-0.25	1.17	0.38	-0.3
t-EGARCH 1 (20)	-1.50	1.03	-1.40	–	-0.87	1.78	-1.77	-1.77	-1.15	-0.77	-0.69	-1.2
Normal GJR 1 (21)	-1.38	1.39	-0.33	0.87	–	1.74	-1.01	-1.01	-0.82	0.25	0.23	-0.54
t-GJR 1 (22)	-2.39	-0.86	-1.99	-1.78	-1.74	–	-2.72	-2.72	-1.90	-1.42	-1.41	-1.98
Normal EWMA 0.94 1 (23)	0.52	2.49	0.75	1.77	1.01	2.72	–	3.26	0.64	1.07	1.03	1.55
t-EWMA 0.94 1 (24)	0.52	2.48	0.75	1.77	1.01	2.72	-3.26	–	0.64	1.06	1.03	1.54
Normal GARCH 2 (25)	-0.44	1.72	0.25	1.15	0.82	1.90	-0.64	-0.64	–	0.85	1.25	-0.17
Normal EGARCH 2 (26)	-1.02	1.14	-1.17	0.77	-0.25	1.42	-1.07	-1.06	-0.85	–	-0.13	-0.66
Normal GJR 2 (27)	-1.17	1.13	-0.38	0.69	-0.23	1.41	-1.03	-1.03	-1.25	0.13	–	-0.63
Normal EWMA 0.94 2 (28)	0.01	1.76	0.30	1.20	0.54	1.98	-1.55	-1.54	0.17	0.66	0.63	–

Table A.3: **Summary Statistics on Margins for 6-month and 12-month WTI Futures**

This table reports statistics for 6-month and 12-month WTI futures margins based on a tolerance setting $\tau_{\text{low}} = \tau_{\text{high}} = 20$ days and for the EWMA and t-GJR MTL models. The model name in column 1 starts with the margin objective (S, T or E) followed by E or G, according to the MTL model (either EWMA or t-GJR GARCH). The same objectives are used for the scheme introduced by Lam *et al.* (2010), and these models have an additional L (e.g. SLE denotes the Lam *et al.* (2010) model with stability objective based on EWMA). The other margin schemes are of Chiu *et al.* (2006), also with two possible MTL models (denoted CE and CG) and, finally the CME. Columns 2 – 4 report the calibrated parameters and subsequent columns provide summary statistics.

	a	b	c	No. of changes	No. of in-creases	No. of de-creases	Avg. days be-tween changes	Avg. margin in-crease (%)	Avg. margin de-crease (%)	Smallest in-crease (%)	Smallest de-crease (%)
					Panel A: WTI (6M)						
SE	0.00	0.11	0.25	21	9	12	92.40	25.49	-20.89	2.82	-13.15
SG	0.08	0.05	0.25	26	10	16	75.08	34.36	-19.34	10.46	-8.59
TE	0.00	0.00	0.34	26	8	18	70.96	24.84	-12.63	0.51	-3.38
TG	0.00	0.00	0.25	27	12	15	72.19	20.22	-17.23	0.63	-7.93
EE	0.00	0.00	0.26	29	9	20	62.89	26.18	-12.82	7.05	-1.48
EG	0.00	0.00	0.25	25	10	15	78.21	22.46	-16.22	0.73	-6.73
SLE	2.63	0.48	–	8	4	4	204.71	67.11	-48.30	56.60	-48.23
SLG	1.56	0.47	–	17	9	8	100.12	62.54	-47.68	47.30	-47.05
TLE	2.48	0.25	–	29	13	16	65.68	35.21	-25.68	25.33	-25.09
TLG	1.56	0.27	–	51	25	26	37.48	36.91	-27.58	27.31	-26.60
ELE	2.43	0.25	–	29	13	16	65.68	35.32	-25.72	25.33	-25.15
ELG	2.21	0.27	–	51	25	26	37.48	36.91	-27.58	27.31	-26.60
CE	–	–	–	76	33	43	25.13	22.79	-15.99	15.31	-15.01
CG	–	–	–	146	64	82	13.10	26.29	-17.15	15.06	-15.05
CME			–	**31**	**14**	**17**	**58.63**	**11.06**	**-11.79**	**2.08**	**-4.76**
					Panel B: WTI (12M)						
SE	0.02	0.11	0.25	25	9	16	76.88	38.95	-20.92	8.15	-11.86
SG	0.03	0.19	0.25	30	17	13	64.45	27.47	-30.06	8.33	-19.20
TE	0.00	0.00	0.40	24	8	16	77.13	21.42	-12.98	0.73	-1.85
TG	0.00	0.00	0.25	28	11	17	69.22	30.96	-18.23	5.65	-5.77
EE	0.00	0.05	0.31	23	8	15	80.05	28.78	-16.37	0.64	-8.09
EG	0.00	0.00	0.25	27	11	16	71.88	30.31	-19.02	5.62	-5.77
SLE	2.66	0.48	–	10	5	5	170.89	67.42	-48.16	52.47	-47.90
SLG	3.01	0.49	–	13	7	6	129.42	66.57	-51.03	50.66	-49.47
TLE	2.47	0.25	–	31	14	17	58.33	37.17	-26.12	25.66	-25.45
TLG	1.60	0.27	–	54	25	29	35.68	42.27	-27.90	27.01	-26.74
ELE	2.69	0.25	–	27	12	15	67.35	36.42	-25.72	25.63	-25.04
ELG	2.35	0.26	–	64	30	34	29.60	40.01	-26.87	26.28	-25.89
CE	–	–	–	83	36	47	22.77	23.12	-15.86	15.31	-15.04
CG	–	–	–	134	54	80	14.23	29.79	-16.47	15.05	-15.04
CME			–	**27**	**11**	**16**	**67.65**	**13.93**	**-12.63**	**5.56**	**-7.14**

Table A.4: **Summary Statistics on Margins for 3-month and 6-month Gold Futures**

This table reports statistics for 3-month and 6-month gold futures margins based on a tolerance setting $\tau_{\text{low}} = \tau_{\text{high}} = 20$ days and for the EWMA and *t*-GJR MTL models. The model name in column 1 starts with the margin objective (S, T or E) followed by E or G, according to the MTL model (either EWMA or *t*-GJR GARCH). The same objectives are used for the scheme introduced by Lam *et al.* (2010), and these models have an additional L (e.g. SLE denotes the Lam *et al.* (2010) model with stability objective based on EWMA). The other margin schemes are of Chiu *et al.* (2006), also with two possible MTL models (denoted CE and CG) and, finally the CME. Columns 2 – 4 report the calibrated parameters and subsequent columns provide summary statistics.

	a	*b*	*c*	No. of changes	No. of in-creases	No. of de-creases	Avg. days be-tween changes	Avg. margin in-crease (%)	Avg. margin de-crease (%)	Smallest in-crease (%)	Smallest de-crease (%)
					Panel A: GC (3M)						
SE	0.05	0.14	0.25	20	7	13	93.63	56.27	-22.43	15.11	-17.33
SG	0.00	0.10	0.25	19	8	11	95.00	37.95	-20.29	2.41	-10.44
TE	0.00	0.00	0.34	26	7	19	68.36	36.63	-11.28	3.02	-3.17
TG	0.00	0.00	0.26	21	7	14	85.65	32.18	-12.03	1.47	-0.38
EE	0.10	0.00	0.46	20	5	15	90.00	45.24	-12.45	16.89	-6.26
EG	0.00	0.00	0.25	21	7	14	85.65	32.50	-12.09	2.05	-0.38
SLE	2.99	0.47	–	9	5	4	167.38	53.81	-47.59	48.42	-47.02
SLG	2.04	0.45	–	14	8	6	135.46	58.91	-46.42	46.87	-45.53
TLE	2.54	0.25	–	38	18	20	50.27	36.62	-25.64	25.34	-25.02
TLG	2.24	0.25	–	40	19	21	46.92	38.96	-26.50	25.36	-25.08
ELE	2.82	0.25	–	38	18	20	50.27	36.83	-25.72	25.34	-25.11
ELG	2.79	0.30	–	24	12	12	74.74	41.79	-30.67	30.18	-30.14
CE	–	–	–	68	28	40	27.84	26.33	-15.74	15.01	-15.02
CG	–	–	–	120	54	66	16.04	24.61	-16.70	15.01	-15.12
CME			**–**	**26**	**11**	**15**	**71.52**	**17.78**	**-10.87**	**5.86**	**-6.25**
					Panel B: GC (6M)						
SE	0.05	0.15	0.25	20	7	13	93.63	57.35	-22.74	15.57	-13.41
SG	0.03	0.12	0.25	19	8	11	94.94	40.58	-21.28	5.09	-13.71
TE	0.00	0.00	0.34	28	8	20	63.33	33.54	-11.17	2.75	-3.20
TG	0.00	0.00	0.27	22	6	16	81.38	45.56	-13.33	14.67	-1.31
EE	0.07	0.00	0.50	18	5	13	100.59	39.08	-12.91	11.39	-4.70
EG	0.00	0.00	0.25	20	6	14	89.95	45.92	-15.24	16.11	-5.84
SLE	3.01	0.47	–	9	5	4	168.62	58.85	-47.32	51.25	-46.98
SLG	1.68	0.46	–	11	6	5	135.40	60.12	-46.81	46.98	-46.02
TLE	2.50	0.25	–	38	18	20	50.27	36.49	-25.55	25.18	-25.05
TLG	2.13	0.25	–	35	16	19	53.76	42.91	-26.79	25.06	-25.05
ELE	2.82	0.27	–	27	13	14	65.23	38.87	-27.49	27.63	-26.98
ELG	2.55	0.27	–	35	17	18	53.79	40.10	-28.57	26.75	-26.77
CE	–	–	–	70	29	41	27.03	26.01	-15.71	15.03	-15.00
CG	–	–	–	111	50	61	17.02	25.50	-17.00	15.17	-15.05
CME			**–**	**26**	**11**	**15**	**71.52**	**17.78**	**-10.87**	**5.86**	**-6.25**

Table A.5: **Summary Statistics on Margins for 3-month and 4-month Euro/Dollar Futures**

This table reports statistics for 3-month and 4-month Euro/Dollar futures margins based on a tolerance setting $\tau_{\text{low}} = \tau_{\text{high}} = 20$ days and for the EWMA and t-GJR MTL models. The model name in column 1 starts with the margin objective (S, T or E) followed by E or G, according to the MTL model (either EWMA or t-GJR GARCH). The same objectives are used for the scheme introduced by Lam *et al.* (2010), and these models have an additional L (e.g. SLE denotes the Lam *et al.* (2010) model with stability objective based on EWMA). The other margin schemes are of Chiu *et al.* (2006), also with two possible MTL models (denoted CE and CG) and, finally the CME. Columns 2 – 4 report the calibrated parameters and subsequent columns provide summary statistics.

	a	b	c	No. of changes	No. of increases	No. of decreases	Avg. days between changes	Avg. margin increase (%)	Avg. margin decrease (%)	Smallest increase (%)	Smallest decrease (%)
					Panel A: EC (3M)						
SE	0.06	0.16	0.25	19	8	11	99.44	24.45	-22.40	15.15	-19.41
SG	0.06	0.13	0.25	23	10	13	81.59	17.59	-17.89	9.74	-11.98
TE	0.00	0.01	0.26	29	11	18	64.04	12.75	-11.51	0.08	-6.09
TG	0.01	0.01	0.25	29	12	17	63.75	9.74	-10.77	1.49	-3.27
EE	0.04	0.02	0.25	28	9	19	66.33	16.58	-11.24	9.13	-2.30
EG	0.00	0.14	0.25	23	12	11	81.14	13.53	-20.01	3.84	-12.31
SLE	2.49	0.44	–	5	2	3	338.25	55.40	-43.85	52.14	-43.69
SLG	3.05	0.45	–	3	1	2	658.00	50.09	-45.33	50.09	-45.14
TLE	2.56	0.25	–	23	10	13	84.73	31.31	-25.54	25.24	-25.00
TLG	2.56	0.25	–	19	8	11	103.50	32.02	-25.59	25.42	-25.02
ELE	2.60	0.25	–	23	10	13	84.73	31.31	-25.54	25.24	-25.00
ELG	2.63	0.25	–	22	10	12	88.71	29.57	-26.06	25.37	-25.28
CE	–	–	–	57	23	34	33.29	23.57	-15.64	15.28	-15.07
CG	–	–	–	52	22	30	36.55	21.94	-15.94	15.22	-15.01
CME			–	**25**	**13**	**12**	**68.96**	**11.77**	**-12.79**	**5.00**	**-6.25**
					Panel B: EC (4M)						
SE	0.06	0.16	0.25	17	7	10	111.88	24.22	-22.50	15.51	-13.73
SG	0.10	0.14	0.25	22	9	13	88.19	21.08	-18.04	10.51	-11.62
TE	0.00	0.00	0.25	30	11	19	62.76	13.18	-11.87	2.97	-4.91
TG	0.00	0.02	0.25	28	11	17	66.11	9.07	-9.94	1.07	-2.12
EE	0.04	0.02	0.25	28	9	19	66.33	16.57	-11.23	9.16	-2.14
EG	0.00	0.13	0.25	23	12	11	81.14	12.19	-18.96	4.84	-12.50
SLE	2.53	0.44	–	7	3	4	281.50	49.04	-44.11	44.03	-43.95
SLG	2.96	0.44	–	3	1	2	658.00	46.42	-44.71	46.42	-44.38
TLE	2.65	0.25	–	23	10	13	84.73	31.56	-25.64	25.01	-25.05
TLG	2.47	0.25	–	17	7	10	116.44	31.21	-25.42	25.44	-25.00
ELE	2.61	0.25	–	23	10	13	84.73	31.56	-25.64	25.01	-25.05
ELG	2.62	0.25	–	20	9	11	97.84	28.81	-25.77	25.37	-25.31
CE	–	–	–	58	24	34	32.70	22.36	-15.57	15.02	-15.05
CG	–	–	–	49	21	28	38.83	20.86	-15.83	15.07	-15.04
CME			–	**25**	**13**	**12**	**68.96**	**11.77**	**-12.79**	**5.00**	**-6.25**

Table A.6: **Summary Statistics on Margins for 3-month and 4-month S&P500 Futures**

This table reports statistics for 3-month and 4-month S&P500 futures margins based on a tolerance setting $\tau_{\text{low}} = \tau_{\text{high}} = 20$ days and for the EWMA and t-GJR MTL models. The model name in column 1 starts with the margin objective (S, T or E) followed by E or G, according to the MTL model (either EWMA or t-GJR GARCH). The same objectives are used for the scheme introduced by Lam *et al.* (2010), and these models have an additional L (e.g. SLE denotes the Lam *et al.* (2010) model with stability objective based on EWMA). The other margin schemes are of Chiu *et al.* (2006), also with two possible MTL models (denoted CE and CG) and, finally the CME. Columns 2 – 4 report the calibrated parameters and subsequent columns provide summary statistics.

	a	b	c	No. of changes	No. of increases	No. of decreases	Avg. days between changes	Avg. margin increase (%)	Avg. margin decrease (%)	Smallest increase (%)	Smallest decrease (%)
Panel A: SP (3M)											
SE	0.02	0.13	0.25	29	12	17	66.64	44.88	-22.79	7.47	-13.58
SG	0.08	0.09	0.25	44	21	23	42.19	62.25	-33.69	8.67	-13.01
TE	0.00	0.05	0.25	31	10	21	62.10	43.09	-16.03	2.70	-5.07
TG	0.00	0.00	0.33	34	15	19	53.39	49.38	-24.99	1.94	-4.70
EE	0.18	0.00	0.49	27	5	22	72.23	72.24	-12.82	17.82	-0.70
EG	0.02	0.00	0.49	24	11	13	76.48	48.50	-27.13	3.82	-8.41
SLE	2.99	0.48	–	9	5	4	222.38	57.25	-48.99	48.78	-48.65
SLG	2.92	0.49	–	35	19	16	52.88	77.78	-49.91	49.50	-48.84
TLE	1.08	0.25	–	34	16	18	57.58	33.83	-25.64	25.20	-25.05
TLG	1.00	0.38	–	91	46	45	20.39	62.53	-38.94	37.88	-37.62
ELE	4.44	0.31	–	17	8	9	118.00	41.79	-31.75	31.90	-31.29
ELG	4.04	0.28	–	149	70	79	12.51	49.36	-29.21	27.77	-27.84
CE	–	–	–	85	38	47	22.69	22.04	-15.82	15.02	-15.01
CG	–	–	–	332	128	204	5.73	35.80	-16.98	15.25	-15.00
CME			**–**	**8**	**5**	**3**	**247.29**	**6.31**	**-10.90**	**3.26**	**-9.09**
Panel B: SP (4M)											
SE	0.02	0.13	0.25	31	12	19	62.70	43.17	-20.93	8.65	-12.86
SG	0.16	0.14	0.25	37	19	18	50.33	72.68	-41.23	16.21	-25.26
TE	0.00	0.08	0.25	33	13	20	58.28	35.89	-17.82	0.60	-7.72
TG	0.00	0.00	0.33	33	15	18	55.06	48.76	-26.10	2.10	-5.54
EE	0.18	0.00	0.49	27	5	22	72.23	71.14	-12.75	17.82	-0.70
EG	0.01	0.00	0.48	24	11	13	76.48	48.22	-26.93	3.88	-8.29
SLE	3.05	0.48	–	9	5	4	222.38	57.29	-49.00	48.81	-48.63
SLG	2.90	0.49	–	35	19	16	52.88	77.98	-49.95	49.45	-49.06
TLE	1.08	0.25	–	34	16	18	57.58	33.83	-25.64	25.25	-25.04
TLG	1.00	0.36	–	88	42	46	21.09	68.36	-37.77	37.10	-36.27
ELE	4.44	0.31	–	17	8	9	118.00	41.91	-31.80	32.02	-31.41
ELG	4.04	0.25	–	168	78	90	11.35	45.74	-27.19	25.42	-25.45
CE	–	–	–	85	38	47	22.69	21.94	-15.78	15.07	-15.00
CG	–	–	–	337	131	206	5.68	35.22	-16.97	15.02	-15.00
CME			**–**	**8**	**5**	**3**	**247.29**	**6.31**	**-10.90**	**3.26**	**-9.09**

Figure A.1: **Comparison of Margins with RMSE Calibration Objective: 3-month WTI Futures**

A comparison of our margin with the CME margin for 3-month WTI futures. Each figure depicts our margin and the upper and lower bounds MTL and $(1+c)$MTL. The MTL model is either the Student t-GJR, with parameters calibrated to the underlying futures returns (upper) or Student t-EWMA with $\lambda = 0.96$ (lower). The calibrated values for the proportional margin decrease a and increase b are reported above each chart. The calibration objective here is RMSE, as explained in the text.

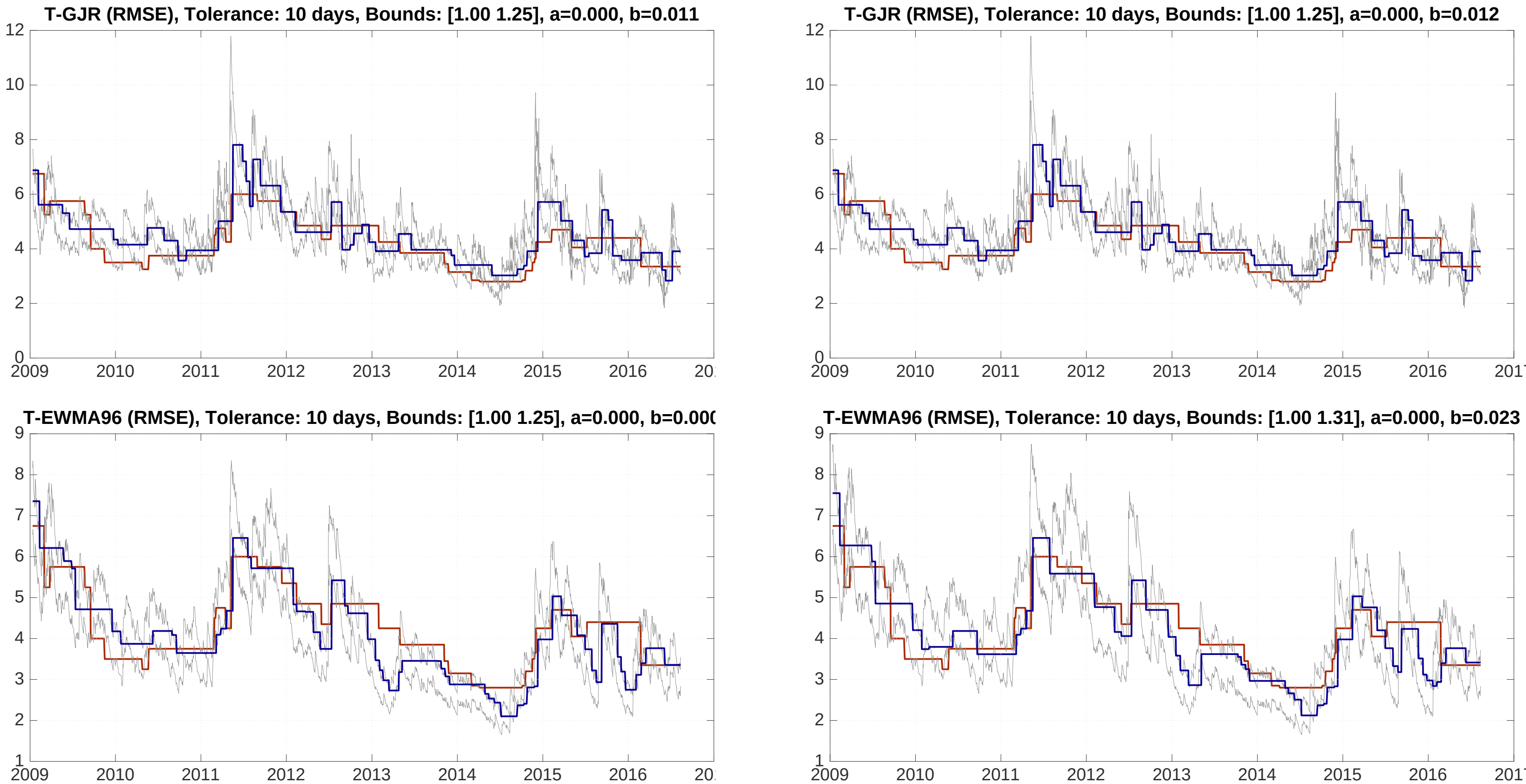


Figure A.2: **Comparison of Margins with Stability Calibration Objective: 3-month WTI Futures**

A comparison of our margin with the CME margin for 3-month WTI futures. Each figure depicts our margin and the upper and lower bounds MTL and $(1+c)$MTL. The MTL model is either the Student t-GJR, with parameters calibrated to the underlying futures returns (upper) or Student t-EWMA with $\lambda = 0.96$ (lower). The calibrated values for the proportional margin decrease a and increase b are reported above each chart. The calibration objective here is Stability, as explained in the text.

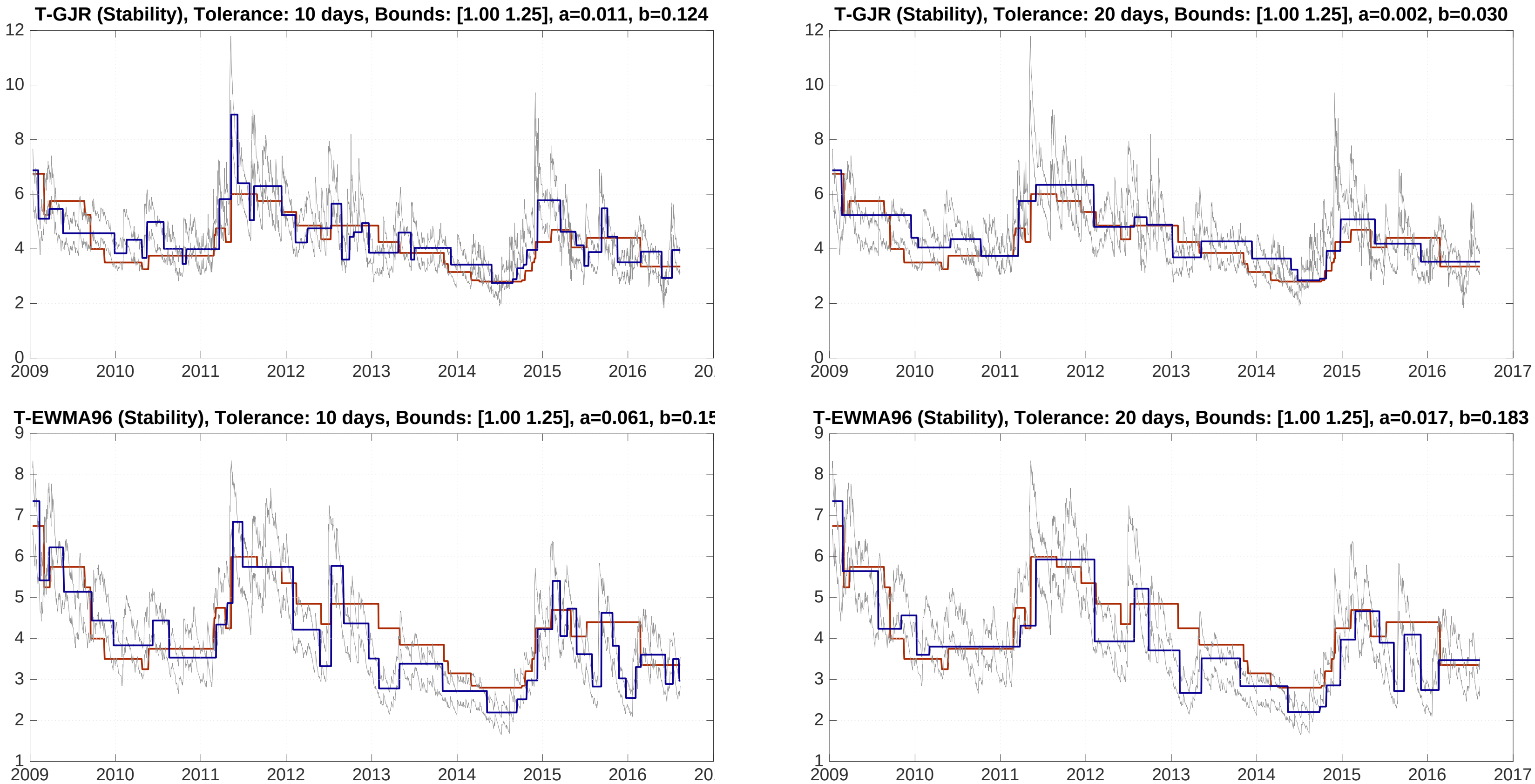


Table A.7: **Summary Statistics on Margins for 1-month and 3-month WTI Futures**

This table reports statistics for 1-month and 3-month WTI futures margins based on a tolerance setting $\tau_{\text{low}} = \tau_{\text{high}} = 20$ days and for the implied volatility based MTL models (see Section 5.3). The model name in column 1 starts with the margin objective (S, T or E) followed by I, indicating that the MTL model is based on implied volatility. The other margin schemes are of Chiu *et al.* (2006) and the CME. Columns 2 –4 report the calibrated parameters and subsequent columns provide summary statistics.

	a	b	c	No. of changes	No. of increases	No. of decreases	Avg. days between changes	Avg. margin increase (%)	Avg. margin decrease (%)	Smallest increase (%)	Smallest decrease (%)
Panel A: WTI (1M)											
SI	0.09	0.08	0.25	16	6	10	105.27	22.96	-15.10	13.51	-8.62
TI	0.00	0.00	0.25	21	7	14	81.80	15.14	-9.67	10.26	-2.01
EI	0.01	0.07	0.25	17	7	10	103.56	15.02	-12.95	8.16	-6.41
SLI	2.54	0.46	–	4	2	2	213.67	55.98	-47.95	54.41	-47.41
TLI	2.67	0.25	–	16	8	8	111.60	29.81	-27.76	25.25	-25.01
ELI	2.46	0.25	–	16	8	8	111.60	29.81	-27.76	25.25	-25.01
CI	–	–	–	51	25	26	36.68	19.75	-16.97	15.09	-15.05
CME			–	**30**	**12**	**18**	**60.66**	**13.53**	**-11.00**	**1.72**	**-4.00**
Panel B: WTI (3M)											
SI	0.07	0.09	0.25	13	4	9	131.58	23.66	-14.30	11.38	-7.92
TI	0.00	0.00	0.25	23	6	17	71.18	17.92	-8.52	4.02	-1.70
EI	0.00	0.12	0.25	17	8	9	103.50	16.44	-17.46	3.21	-9.99
SLI	3.01	0.47	–	1	0	1	–	0.00	-48.51	0.00	-48.51
TLI	2.66	0.25	–	21	10	11	83.70	31.36	-26.42	25.47	-25.12
ELI	2.29	0.25	–	21	10	11	83.70	31.36	-26.42	25.47	-25.12
CI	–	–	–	51	25	26	36.70	18.88	-16.85	15.02	-15.04
CME			–	**31**	**13**	**18**	**58.63**	**12.81**	**-11.42**	**1.79**	**-1.75**